IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re application of:	Confirmation No.: 9891
SNYDER *et al.*	Art Unit: 1639
Application No.: 09/849,781	Examiner: WESSENDORF, Teresa D.
Filed: May 4, 2001	Atty. Docket: 2681.0030002/RWE/JKM
For: **Protein Chips for High Throughput Screening of Protein Activity**	

Supplemental Declaration of Barry Schweitzer, Ph.D. Under 37 C.F.R. § 1.132

Attn: Mail Stop Amendment

Commissioner for Patents
PO Box 1450
Alexandria, VA 22313-1450

Sir:

I, Dr. Barry Schweitzer, residing at 459 Maple Avenue, Cheshire, CT, USA, do hereby declare and state as follows:

1.   I am currently employed by Life Technologies Inc. ("LTI"), a licensee of the above-captioned application.   I hold the position of Director of Integrated Technologies, Molecular Biology Systems Division. A copy of my *curriculum vitae* is attached hereto as Exhibit 1. I received my Ph.D. degree in Pharmacology from Yale University. As indicated by my attached *curriculum vitae*, I have published many papers relating to protein microarrays. Based on my education and experience, I am an expert in the fields of genomics, molecular genetics, and proteomics (including protein microarraying).

2.   I have read and understand the following documents:

- U.S. Application No. 09/849,781 ("the '781 application"; Exhibit 2)
- Pending claims (Exhibit 3)
- Office Action mailed July 7, 2009 ("the first Office Action"; Exhibit 4)
- Office Action mailed July 9, 2010 ("the second Office Action"; Exhibit 5)

- Declaration entitled "Declaration of Barry Schweitzer, Ph.D., Under 37 C.F.R. §1.132" with Exhibits A-D, dated March 2, 2010 ("my first Declaration"; Exhibit 6)
- Hanks et al., FASEB J. 9:576-596 (1995) ("Hanks1"; Exhibit 7)
- Hanks et al., Science 241:42-52 (1988) ("Hanks2"; Exhibit 8)
- Bold et al., Mol. Cell. Biol. 5(11) 331-338 (1985) ("Bold"; Exhibit 9)
- Yaciuk et al., Mol. Cell. Biol. 6(8):2807-2819 (1986) ("Yaciuk"; Exhibit 10)
- Sadowski et al., Oncogene 1:181-191 (1987) ("Sadowski"; Exhibit 11)
- Hubbard et al., Nature 372 (6508):746-754 (1994) (Hubbard; Exhibit 12)
- Hanks et al., Genome Biology 4:111-111.7 (2003) ("Hanks3"; Exhibit 13)
- Hunter and Plowman TIBS 22:18-22 (1997) ("Hunter and Plowman"; Exhibit 14)
- Plowman et al., Proc. Natl. Acad. Sci. 96:13603-13610 (1999) ("Plowman"; Exhibit 15)
- Morrison et al., J. Biol. Chem. 150(2):F57-F62 (2000) ("Morrison"; Exhibit 16)
- Manning et al., Science 298:1912-1934 (2002) ("Manning"; Exhibit 17)
- Shaw et al., Drug Discovery and Development (2005) ("Shaw"; Exhibit 18)
- Zhu et al., Nature Genetics 26:283-289 (2000) ("Zhu"; Exhibit 19)
- Tleugabulova et al., J. Chrom. B. 720:153-163 (1998) ("Tleugabulova"; Exhibit 20)
- Groll et al., J. Am. Chem. Soc. 126:4234-4239 (2004) ("Groll"; Exhibit 21)
- Bussow et al., Nucleic Acids Res. 26(21):5007-5008 (1998) ("Bussow1"; Exhibit 22)
- Bussow et al., Genomics 65:11-8 (2000) ("Bussow2"; Exhibit 23)

3.   In formulating my opinions set forth in this declaration, I have considered the viewpoint of a scientist of ordinary skill in the field of proteomics, as of May 4, 2001, the filing date of the '781 application. I understand that a scientist of ordinary skill in the field of proteomics is a hypothetical scientist who thinks along conventional wisdom in the field of proteomics, and is a scientist of ordinary creativity who does not seek to innovate. A scientist of ordinary skill in the field of proteomics would have had general knowledge of the scientific literature concerning proteomics (including protein microarraying) technology that was available by May 4, 2001, including knowledge about the identification and characterization of protein kinases and also experimental techniques and results available in the art. A scientist of ordinary skill in the field of

proteomics would have a scientific background and hold a Master's degree or Ph.D. in the biological and/or physical sciences (e.g., pharmacology), and have substantial familiarity, training, and experience with proteomics (including protein microarraying). It is my understanding that a scientist of ordinary skill in the field of proteomics is a hypothetical scientist, whereas an expert in the field of proteomics is an actual scientist. I believe that a scientist of ordinary skill in the field of proteomics would agree with my opinions expressed in this declaration.

4.   As an expert in the field of proteomics technology since before 2001, I am qualified to provide an opinion as to what a scientist of ordinary skill in the field of proteomics would have known and concluded as of May 4, 2001.

5.   I understand that claim 1 of the '781 application is directed to a positionally addressable array comprising a plurality of different substances on a solid support, with each different substance being at a different position on the solid support, wherein the density of the different substances on the solid support is at least 100 different substances per cm^2, and wherein the plurality of different substances comprises 61 purified active kinases (or functional kinase domains thereof) of a mammal, 61 purified active kinases (or functional kinase domains thereof) of a yeast, or 61 purified active kinases (or functional kinase domains thereof) of a *Drosophila*.[1]

6.   I also understand that claim 1 and the claims that depend from claim 1, have been rejected by the Patent Office for allegedly failing to comply with requirements for written description and enablement.

7.   I further understand the Patent Office has taken the position that the '781 application, which includes the exemplary disclosure of arrays containing 111 distinct

purified active yeast protein kinases, fails to provide an adequate description that a scientist of ordinary skill would have reasonably believed to be generally applicable to distinguishing protein kinases and to predictably arraying at least 61 purified active protein kinases (or fragments containing a functional kinase domain) of any organism, such as a yeast, a *Drosophila* or a mammal.[2] The Patent Office has also argued that since biotechnology is unpredictable, the '781 application does not provide sufficient disclosure to teach a scientist of ordinary skill how to make and use the full scope of the claimed protein arrays containing at least 61 purified active protein kinases (or fragments containing a functional kinase domain) from an organism, such as a mammal, a yeast, or a *Drosophila*, without having to undertake inordinate experimentation.[3]

8.   I have reviewed the Patent Office's statements relating to its bases for making these rejections, and I have concluded that the rejections are founded on flawed and/or unsubstantiated reasoning that lead to incorrect conclusions. More particularly, the Patent Office's rejections relating to written description and enablement are founded on conclusory arguments that overlook the disclosure and guidance provided by the '781 application and the extensive knowledge of scientists of ordinary skill in the field of proteomics on May 4, 2001. For at least the reasons discussed below, it is my opinion that the disclosure of the '781 application would indeed have reasonably been understood by a scientist of ordinary skill in the field of proteomics on May 4, 2001 to encompass and provide ample disclosure of the claimed protein kinase arrays, which contain at least 61 purified active protein kinases (including fragments containing a functional kinase

---

[1] See Pending claims (Exhibit 3).

[2] Second Office Action at pages 3-10 (Exhibit 5).

[3] Second Office Action at pages 11-20 (Exhibit 5).

domain) of a yeast or another organism, such as a *Drosophila* or a mammal (e.g., a human), as well as disclosure and guidance that would have been reasonably expected to be generally applicable to making and using these arrays. It is also my opinion that a scientist of ordinary skill in the field of proteomics on May 4, 2001, enlightened by the disclosure and guidance of the '781 application, would have reasonably expected to be able to, and indeed would have been able to, make and use the protein kinase arrays that fall within the scope of claim 1[4] without having to undertake more than routine experimentation. A discussion relating to the written description and enablement rejections are addressed in turn below.

Disclosure of the '781 application

9.   Well before May 4, 2001, protein kinases were known to represent a large superfamily made up of hundreds of proteins that were assigned to this superfamily by virtue of their containing a kinase domain.[5] For example, by May *1995*, "about 200 different superfamily members (products of distinct paralogous genes) had been recognized from mammalian sources alone!"[6] Moreover, protein kinase domains were known to consist of discrete polypeptide regions of approximately 250-300 amino acid residues and to contain characteristic patterns of conserved residues, including twelve invariant or nearly invariant residues.[7] The conserved patterns and invariant residues in

---

[4] See Pending claims (Exhibit 3).

[5] See e.g., the '781 application at page 9, lines 1-8; page 31, lines 8-17; and Figures 5a and 5b (Exhibit 2). See also, e.g., Hanks1 (Exhibit 7) referred to on page 31, line 15, and page 41, lines 11-12 of the '781 application (Exhibit 2), which has been incorporated by reference (page 46, lines 14-17).

[6] Hanks1, at page 576, column 2, first paragraph (Exhibit 7).

[7] See e.g., Hanks2 at page 42 and pages 45-46 (Exhibit 8); Bold (Exhibit 9); Yaciuk (Exhibit 10); and Sadowski (Exhibit 11).

these kinase domains were known to play essential roles in enzyme function, as

corroborated by the crystal structure of several protein kinase superfamily members.[8] By

May 4, 2001, alignments and characterization studies of scores of protein kinases and

their protein kinase domains had identified differences and alterations in kinase domain

sequences that retained, abolished, and distinguished kinase activity.[9] These studies also

identified and characterized distinct conserved sequence motifs in a relative handful of

proteins that had been reported to have kinase activity, but that lacked a conventional

kinase domain.[10] Moreover, since from well before May 4, 2001 to present, sequence

analysis has routinely been applied to reliably identify and distinguish protein kinases.[11]

Thus, on May 4, 2001, a scientist of ordinary skill in the field of proteomics would have

readily comprehended that protein kinases contain a highly conserved catalytic kinase

domain for which there was an art-recognized correlation between primary amino acid

sequence (i.e., structure) and kinase activity (i.e., function), and that protein kinases were

---

[8] See e.g., Hanks1 at page 576-577, and page 588, col.1 to page 592 col. 2, which discuss the crystal structures of the protein kinases PKA-Calpha, Erk2, twichin kinase, casein kinase I, Cdk2, and the insulin receptor; see also note added in proof at page 592, col. 2 (Exhibit 7); see also Hubbard (Exhibit 12).

[9] See for example, initial screening of the yeast genome by Hunter and Plowman at page 14 (Exhibit 14), which has been incorporated by reference in the '781 application (page 46, lines 14-17 (Exhibit 2)).

[10] See, e.g., Plowman at page 13609 and Table 1 (Exhibit 15) which has been incorporated by reference in the '781 application (page 46, lines 14-17 (Exhibit 2)); Hanks3 at page 111.2, col. 1 first full paragraph (Exhibit 13); and Morrison at page F59 (Exhibit 11).

[11] For example, sequence analysis was relied to identify protein kinase members in the initial screening of the yeast genome by Hunter and Plowman (Exhibit 14; referred to on page 34, line 2 of the '781 application (Exhibit 2)), the *Drosophila* genome by Morrison (Exhibit 16), and the human genome by Manning (Exhibit 17), were all initially identified using homology based analysis. See also Plowman at pages 13604-13608 and Table 11 (Exhibit 15; referred to on page 27, lines 27-29, and incorporated by reference at page 46, lines 14-17 of the '781 application (Exhibit 2)).

- 7 -

SNYDER *et al.*
Appl. No. 09/849,781

and could be, reliably recognized and distinguished based on the sequence of their kinase domain.

10.   Against this backdrop, the '781 application refers to a screen of the yeast genome conducted by Hunter and Plowman that identified 122 open reading frames predicted to encode the protein kinases of the yeast genome (i.e., the yeast kinome).[12] Example 1, at pages 27-41 of the '781 application, discloses the manufacture and screening of arrays containing 119 of the 122 yeast protein kinases identified by Hunter and Plowman and reports that "most (i.e., 93% [111 kinases]) kinases" on the arrays exhibit protein kinase activity.[13] Thus, Example 1 of the '781 application confirms that on May 4, 2001, protein kinases were readily and reliably recognized and distinguished based on the primary amino acid sequence of their catalytic kinase domain.

11.   The '781 application describes arrays of protein kinases and fragments containing a functional kinase domain from a yeast and other organisms, including a mammal and a *Drosophila*.[14] The '781 application also discloses the production of 17 protein chip arrays containing 111 distinct purified active yeast protein kinases (i.e., 93% of the yeast kinome) and states that these arrays are intended to be exemplary and non-limiting.[15] More than 90% (i.e., 111/119) of the protein kinases on the arrays disclosed in Example 1 are reported to display kinase activity.[16] The active arrayed proteins

---

[12] '781 application at page 27, line 32 to page 28, line 5; and page 31, lines 6-17 (Exhibit 2); Hunter and Plowman (Exhibit 14), which has been incorporated by reference in the '781 application (page 46, lines 14-17 (Exhibit 3)).

[13] See '781 application at page 8, lines 26-30; and page 33, line 34-35; and Figure 4a (Exhibit 2).

[14] See e.g., '781 application at page 11, second full paragraph (Exhibit 2).

[15] See e.g., '781 application at page 33, lines 14-36 (Exhibit 2).

[16] See e.g., '781 application at page 28, lines 9-14 (Exhibit 2).

include 18 of the 24 previously unstudied yeast protein kinases, and unconventional

protein kinases such as, histidine kinases (Sln1, Yil042c) and phospholipid kinases (e.g.,

Mec1).[17] Example 1 also discloses that 27 of the arrayed kinases display tyrosine kinase

activity and are able to phosphorylate the substrate poly(Tyr-Glu).[18] Several of the

kinase-substrate activities disclosed in Example 1 are reported to correspond to known

kinase-substrate relationships and the '781 states that similarly, the other substrates

identified in Example 1 are likely to be *bona fide* substrates for their identified

counterpart protein kinase(s) *in vivo*.[19]

12.   In view of the extensive disclosure and teaching of the '781 application

and the art-recognized correlation between the primary sequence and activity of the

kinase domain, a scientist of ordinary skill in the field of proteomics reading the '781

application on May 4, 2001, would readily have been able to recognize and distinguish

protein kinases and would have reasonably concluded that the '781 application provides

ample description of arrays containing at least 61 purified active protein kinases

(including fragments having a functional kinase domain) from a yeast, or another

organism, such as a *Drosophila,* or a mammal (including a human). In particular, a

scientist of ordinary skill, enlightened by the disclosure and teaching of the '781

application, would have readily understood that the disclosure and teaching in the

application applies to the production of equally successful arrays containing active

protein kinases from other organisms (e.g., a *Drosophila* or a mammal) that could readily

be recognized and distinguished from other proteins and that could be arrayed as active

---

[17] '781 application at page 34, lines 1-4 (Exhibit 2).

[18] '781 application at page 34, lines 27-36 (Exhibit 2).

[19] '781 application at page 36, lines 4-9 (Exhibit 2).

proteins according to the methods and techniques disclosed in the '781 application, irrespective of whether the arrayed kinases were well characterized or uncharacterized, or whether the kinases were then known or yet to be recognized. For at least these reasons, it is my opinion that a scientist of ordinary skill in the field of proteomics on May 4, 2001, reading the '781 application, would have reasonably concluded that the application amply describes the claimed arrays which contain at least 61 purified active kinases (and fragments containing a functional kinase domain) of a yeast, a *Drosophila* or a mammal (e.g., a human).

13.   In the second Office Action, the Patent Office states that the disclosure of the '781 application is limited to that of "a single species" of the claimed genus.[20] In response, I point out that the '781 application discloses the production of 17 arrays containing 111 distinct and diverse purified active yeast protein kinases - well over the lower limit of 61 active kinases recited in the pending claims. Additionally, the high percentage of the large number of yeast protein kinases that display kinase activity on the arrays prepared according to the methods and techniques disclosed in the '781 application (i.e., almost every arrayed protein corresponding to almost every kinase in the yeast kinome) provide compelling support that these methods and techniques could be applied to routinely and predictably produce arrays containing at least 61 active protein kinases (and fragments containing a functional kinase domain) from an organism, such as a yeast, a *Drosophila*, or a mammal. Moreover, the disclosure in Example 1 of the production and use of 17 arrays that each display 111 active yeast protein provides further support that would lead a scientist of ordinary skill in the field of proteomics to

---

[20] The Second Office Action, paragraph spanning pages 4-5 (Exhibit 5)

reasonably conclude that the techniques and methodology described in the '781 application are reproducible, generally applicable to actively arraying large numbers of active diverse protein kinases on a single array, and can be routinely used or adapted to produce and use arrays containing at least 61 active protein kinases (and fragments containing a functional kinase domain) from, for example, a yeast, a human or another mammal, or a *Drosophila*.

14.   Additionally, it is significant to note that many of the statements made by the Patent Office in support of the written description rejection are unsubstantiated and/or non-persuasive in view of the state of the art and the teaching and disclosure of the '781 application. In particular, the Patent Office speculates that "a skilled artisan recognizes that one cannot rule out the possibility that" (a) it might be difficult to array poorly characterized protein kinase family members without denaturing them; (b) kinases other than the desired enzyme can contaminate the purification preparations; and (c) the kind and type of substrate may also be a factor that will influence the activity of an arrayed protein kinase.[21]

15.   In response, I point out that, as would have been immediately apparent to a scientist of ordinary skill reading the '781 application, the presence of poorly characterized protein kinases, the possibility of contamination, and the lack of knowledge relating to substrate identity did not prevent the successful production and use of the arrays exemplified in Example 1. In particular, the '781 application, at page 27, lines 34-35 and page 34, lines 1-4, discloses that 75% (i.e., 18/24) of the arrayed proteins that had not been previously studied (i.e., were "poorly characterized")

---

[21] Second Office Action at page 5, quoting statement in the '781 application at page 36, line 10 (Exhibit 4).

displayed kinase activity and phosphorylated one or more substrates. With respect to the

potential for sample contamination, the '781 application at page 35, lines 29-33, states

that "[o]ne concern with these studies is that it is possible that kinases other than the

desired enzyme are contaminating our preparations. Although this cannot be rigorously

ruled out, analysis of five of our samples by Coomasie staining and immunoblot staining

with anti-GST does not reveal any detectable bands in our preparation that are not GST

fusions (see methods)." Therefore, contamination would not appear to have prevented

the successful production of the arrays containing active protein kinases in Example 1 of

the '781 application. Lastly, with respect to substrate identity, the high percentage of

arrayed protein kinases that demonstrate kinase activity in Example 1 (i.e., 93%),

indicates that the kind and type of substrate did not prevent the successful manufacture

and use of arrays containing purified active kinases according to the methods disclosed

in the '781 application. [22] Accordingly, poorly characterized protein kinases,

contamination, and substrate anonymity did not prevent the successful production and

use of the 11 arrays containing the purified active protein kinases disclosed in Example

1. Moreover, in view of the detailed disclosure and teaching of the '781 application and

the high level of knowledge and skill in the field of proteomics on May 4, 2001, a

scientist of ordinary skill would have reasonably expected that the reagents,

methodology, and techniques disclosed and taught in the '781 application could be

routinely, reliably, and predictably applied to produce protein arrays containing at least

61 purified active proteins kinases from yeast and other organisms, such as, a

Drosophila, or a human or other mammal.

---

[22] '781 application at page 34, lines 27-36; and page 36, lines 4-9 (Exhibit 2).

16.   The Patent Office additionally emphasizes the statement in the '781 application that "although most of the kinases were active in [our] assays, several were not."[23]   In response, I point out that the Patent Office's choice to emphasize this statement reflects its failure to appreciate the full extent of the disclosure and teaching of the '781 application in its entirety, as it would have been understood by a scientist of ordinary skill in the field of proteomics on May 4, 2001. In particular, this misplaced emphasis overlooks the fact that the '781 application discloses the production and use of 17 arrays on which almost every protein kinase in the kinome of an organism (i.e., yeast) displays kinase activity. The high percentage (i.e., greater than 90%) of the large number of distinct and diverse arrayed protein kinases (i.e., 121) that are active on the kinase arrays (i.e., 111) described in Example 1 would have been viewed on May 4, 2001 by a scientist of ordinary skill in the field of proteomics to be a highly successful experiment that represented a substantial technological advancement in protein arraying.

17.   Additionally, in the second Office Action at pages 15-16, the Patent Office relies on statements made in Shaw[24] to support the position that "proteins have proven to be much trickier to work with in array format than their genomic counterparts" and that protein arraying is unpredictable due to issues such as stability, the protein arraying technique utilized, and non-specific binding. In response, while I agree that producing protein arrays might generally be viewed as challenging, particularly, when compared to producing DNA microarrays, I *disagree* that this viewpoint would have prevented a scientist of ordinary skill in art of proteomics on May 4, 2001, enlightened by the disclosure and guidance provided by the '781 application, from reasonably

---

[23] Second Office Action at page 15 (Exhibit 5).

[24] Shaw (Exhibit 18).

- 13 -

SNYDER *et al.*
Appl. No. 09/849,781

concluding that the '781 application adequately describes the making and using of protein arrays containing at least 61 active protein kinases from yeast or another organism, including a *Drosophila* or a mammal, such as a human (including fragments having a functional kinase domain of these protein kinases). In particular, the high percentage (i.e., greater than 90%) of the large number of distinct protein kinases that are active on the arrays disclosed in Example 1 (i.e., 111) indicates that protein stability, choice of immobilization technique and non-specific binding did not prevent the successful arraying and use of protein kinase arrays prepared according to the disclosure and teaching of the '781 application. Furthermore, in view of the known correlation between the primary sequence and activity of the kinase domain and the general applicability of the methods disclosed and taught in the '781 application for recombinantly expressing, purifying, and arraying a large number of distinct and diverse active protein kinases from other organisms, a scientist of ordinary skill in the field of proteomics would have reasonably expected that the methods, techniques and reagents disclosed in the '781 application could be used to produce equally successful arrays containing at least 61 purified active protein kinases (or fragments containing a functional kinase domain) of yeast or another organism, such as a *Drosophila* or a mammal (e.g., a human).

Enablement Rejections of the '781 application

18.   As discussed above, prior to May 4, 2001, there was an art-recognized correlation between the primary sequence (i.e., structure) and kinase activity (i.e., function) of the catalytic domain of protein kinases. Kinase domains were known to represent discreet regions containing characteristic patterns of conserved and invariant or nearly invariant amino acid residues that play essential roles in conferring kinase

activity. Well before May 4, 2001, scientists of ordinary skill in the field of proteomics relied on this known correlation between the sequence and activity of kinase domains to reliably recognize and distinguish protein kinases using primary sequence analyses.[25] The use of sequence analysis to identify protein kinases had been validated through numerous activity studies and is further corroborated by the disclosure in Example 1 of the '781 application, which demonstrates that almost every yeast protein kinase that was identified by Hunter and Plowman based on primary sequence analysis, displays kinase activity. Accordingly, a scientist of ordinary skill in the field of proteomics on May 4, 2001 would have reasonably expected that protein kinases and fragments containing a functional kinase domain (including fully characterized protein kinases, poorly characterized protein kinases, or proteins predicted to have kinase activity based on deduced polypeptide sequence) would display kinase activity.

19   The '781 application discloses the recombinant production of chimeric fusion proteins containing a tag (e.g., glutathione-S-transferase (GST)) fused to a large number of protein kinases (e.g., corresponding to almost every protein kinase in the kinome of an organism (e.g., *Saccharomyces cerevisiae*)); the use of a reagent having affinity for this tag component (e.g., the affinity of glutathione for the GST tag) to rapidly and efficiently purify the chimeric kinase proteins from host cell lysates at low temperatures (4°C) and under non-denaturing conditions; the design, manufacture and optimization of solid supports and linking agents; and the immobilization and arraying of

---

[25] For example, sequence analysis was relied to identify protein kinase members in the initial screening of the yeast genome by Hunter and Plowman (Exhibit 14; referred to on page 34, line 2 of the '781 application (Exhibit 2)), the *Drosophila* genome by Morrison (Exhibit 16), and the human genome by Manning (Exhibit 17), were all initially identified using homology based analysis. See also Plowman at pages 13604-

the chimeric protein kinases onto the solid support in a manner so as to retain protein

kinase activity, as well as the subsequent assaying of the arrayed protein kinases for

kinase activity.[26]   In Example 1, the methods and teachings disclosed in the '781

application are applied to produce 17 exemplary[27] arrays, each containing 111 distinct

purified active yeast protein kinases that represent almost every protein in the yeast

kinome. More than 90% of the proteins arrayed in Example 1 display kinase activity,[28]

including 18 uncharacterized yeast protein kinases, and unconventional kinases such as,

histidine kinases and phospholipid kinases.[29]   Example 1 also discloses that 27 of the

arrayed kinases display tyrosine kinase activity.[30]   The '781 application additionally

reports that several of the kinase-substrate relationships reported in Example 1,

correspond to known phosphorylation interactions *in vivo*.[31]

        20.   In view of the extensive disclosure, teaching and guidance of the '781

application and the high level of knowledge and skill in the field of proteomics on May

4, 2001, a scientist of ordinary skill, having read the '781 application, would have

reasonably concluded that the disclosure and teaching of the '781 application could be

routinely applied or modified to produce arrays containing at least 61 purified active

protein kinases (including fragments having a functional kinase domain) from an

---

13608 and Table 11 (Exhibit 15; referred to on page 27, lines 27-29 of the '781
application (Exhibit 2)).

   [26] See e.g., '781 application at page 26, line 25, through page 27, line 19; and
page 28, line 3 to page 37, line 34 (Exhibit 2).

   [27] '781 application at page 33, lines 14-15 (Exhibit 2).

   [28] '781 application at page 28, lines 12-14 (Exhibit 2).

   [29] '781 application at page 34, lines 1-4 (Exhibit 2).

   [30] '781 application at page 34, lines 27-36 (Exhibit 2).

   [31] '781 application, at page 36, lines 4-9 (Exhibit 2).

organism, such as from a mammal, a yeast, or a *Drosophila*. Moreover, in view of the

disclosure and teaching of the '781 application, a scientist of ordinary skill would have

reasonably come to this same conclusions irrespective of whether the arrayed protein

kinases are: (a) from a yeast, a *Drosophila*, or from a human or any other mammal; (b)

fully characterized, poorly characterized; or yet to be identified; or (c) full-length, a

fragment containing a functional kinase domain, or a polypeptide predicted to have

protein kinase activity based on its deduced amino acid sequence.

   21.   On pages 12-13 of the second Office Action, the Patent Office states:

> [i]n a highly unpredictable art, as biotechnology, where one cannot
> predict whether one species would be predictive to the huge scope of the
> claims, one cannot make a priori statement without any experimental
> studies. Factors such as the compatibility of the array with the substrates
> and compounds disposed therein, the compounds (kinases) itself and other
> unpredictable variables can affect the active form of any kinase. Thus,
> one cannot predict from a single species its correspondence or
> extrapolation to the genus, as claimed.

In response, I point out that this statement reflects the Patent Office's failure to fully

appreciate the state of the knowledge relating to protein kinases on May 4, 2001, the

level of skill of scientists of ordinary skill on this date, and the extent to which these

scientists would have understood the disclosure, teaching, and guidance of the '781

application to be generally applicable to, and predictive of, the ability to routinely make

and use the claimed arrays containing at least 61 purified active proteins kinases of a

yeast, a *Drosophila* or a mammal (including fragments having a functional kinase

domain).

   22.   In particular, the '781 application discloses methods and techniques that

are demonstrated to successfully produce protein arrays containing a surprisingly high

percentage (i.e., greater than 90%) of active, distinct and diverse protein kinases

(representing "nearly all" of the protein kinases in the yeast kinome).[32] Given the immediately apparent general applicability of the methods disclosed in the '781 application for recombinantly expressing, purifying, and arraying a large number of active protein kinases, a scientist of ordinary skill in the field of proteomics would have reasonably expected, and correctly so, that the methods and techniques disclosed in the '781 application could be used or routinely modified to produce equally successful arrays containing at least 61 purified active protein kinases (or fragments containing a functional kinase domain) from an organism, such as a *Drosophila,* a yeast, or a human or another mammal.

23.   Moreover, the Patent Office's speculation that factors such as, substrate and array compatibility, distinctions between kinases, and "other unpredictable variables" make the manufacture of the claimed protein kinase arrays unpredictable is baseless in view of the disclosure and teaching of the '781 application, which includes the successful production of the active kinase protein arrays reported in Example 1. In particular, as would be immediately apparent to a scientist of ordinary skill in the field of proteomics on May 4, 2001, substrate and array compatibility, distinctions between kinases, and "other unpredictable variables" did *not* prevent the successful arraying of the purified active yeast protein kinases using the methods and techniques disclosed in the '781 application. Moreover, in my opinion, the Patent Office has provided no reasonable or compelling basis that would have led a scientist of ordinary skill in the field of proteomics on May 4, 2001, enlightened by the disclosure and teaching of the '781 application, to disregard or doubt that the generally applicable disclosure and

---

[32] '781 application at page 28, lines 3-14 (Exhibit 2).

teachings of the '781 application could be equally successfully applied to array purified

active protein kinases from yeast or from other organisms such as a *Drosophila*, or a

mammal.   In particular, I point out that the Patent Office's conclusory and unclearly

articulated statement that "the fact remains that purification the technique [sic.] and other

experimental conditions/steps for yeast would be different from any type of mammals"[33]

is inconsequential and unsubstantiated in view of the advanced state of the art, the high

level of skill in the field of proteomics on May 4, 2001, and the detailed disclosure and

teaching of the '781 application discussed herein.   For example, the specification

discloses that the protein kinases can be recombinantly expressed as fusion proteins

containing tags (e.g., glutathione-S-transferase (GST)) and that these fusion proteins can

be rapidly purified and arrayed using reagents having affinity for such tags (e.g.,

glutathione).[34.] As would be immediately apparent to a scientist of ordinary skill, the use

of the approaches disclosed in the '781 application for recombinantly expressing and

purifying tagged protein kinases using affinity reagents would be expected to be

generally applicable to recombinantly expressing, purifying, and arraying protein kinases

from any organism.   Moreover, I disagree with the Patent Office and in my opinion, a

scientist of ordinary skill, reading the disclosure of the '781 application, particularly

Example 1, would have reasonably concluded that the disclosed methods and techniques

for producing and arraying large numbers of active purified protein kinases, including for

example, methods and techniques for recombinantly producing, purifying and arraying

active protein kinases, would be expected to be equally applicable to, and would require

---

[33] The Second Office Action at page 16, first full paragraph (Exhibit 5)

[34] See, e.g., '781 application at page 17, lines 12-17; and page 32, lines 8-22
(Exhibit 2).

at most routine modification for, successfully producing arrays containing active purified

protein kinases from yeast and other organisms, including for example, a mammal (e.g.,

a human), or a *Drosophila*.

24.   As I have previously testified, when I first read Dr. Michael Snyder's

journal publication corresponding to the '781 application (i.e., Zhu; Exhibit 19) on or

around its publication date, I was surprised by the publication's report of the successful

production of arrays containing such a large number (i.e., 119) of different purified

active protein kinases for which such a high percentage of the arrayed proteins displayed

kinase activity (i.e., 94% (112/119)). Prior to this work, it was generally understood that

the technical limitations associated with preparing and arraying large numbers of

proteins on a single array typically led to protein denaturation and conformational

changes that resulted in a significant loss of protein activity among the arrayed

proteins.[35]

25.   It was only after having been enlightened by the disclosure, teaching and

guidance provided in the '781 application that a scientist of ordinary skill in the field of

proteomics would have been able to routinely produce the claimed arrays containing at

least 61 purified active protein kinases, or fragments containing a functional protein

kinase domain. More particularly, as discussed herein, the '781 application discloses

techniques and methods for recombinantly producing and rapidly purifying and arraying

high densities of large numbers of active protein kinases, as well as the use of the arrays

---

[35] See, e.g., Tleugabulova (Exhibit 20), Groll (Exhibit 21); Bussow1 (Exhibit
22); and Bussow2 (Exhibit 23).

prepared according to the teaching of the '781 application in assaying for kinase activity.[36]

26.   A scientist of ordinary skill in the field of proteomics, enlightened by the disclosure and teaching of the '781 application on May 4, 2001, would have reasonably concluded that, as disclosed in the '781 application, the disclosed methods and techniques and the exemplified production and use of arrays containing purified active yeast protein kinases, could be routinely applied to make and use equally successful arrays containing at least 61 purified active protein kinases or fragments containing function kinase domains, from a yeast, a *Drosophila,* or a mammal (including a human), without having to undertake inordinate or excessive experimentation.

27.   As I have also previously testified, after reading Zhu (Exhibit 19), I joined Protometrix, Inc. (Protometrix), the licensee of Dr. Snyder's protein array technology and subsequent to my joining Protometrix, researchers at the company relied on the information set forth in the '781 application and the known homologies between human and yeast kinases to identify genes encoding kinases and kinase function domains and to successfully manufacture human protein arrays on which at least 61 purified protein kinases or fragments containing kinase domains are active as demonstrated by catalytic activity. It is my belief and understanding that the Protometrix researchers succeeded in making these arrays by expressing proteins in a baculovirus expression system and by relying on, and making at most minor conventional adaptations to the teaching of the '781 application.[37] Thus, no more than routine adaptation of the teaching of the '781

---

[36] See, e.g., '781 application at page 26, line 25, through page 27, line 19; and page 28, line 3 to line 22.

[37] See, e.g., the description of the methods applied by the Protometrix researchers provided in Paragraphs 10-13 first Declaration at (Exhibit 6).

application was applied by the researchers at Protometrix to identify human protein

kinases and their functional domains, and make and use positionally addressable arrays

comprising different proteins on a solid support, with each different protein being at a

different position on the solid support, wherein the density of the different proteins on

the solid support was at least 100 different proteins per cm^2, and contained at least 61

purified active human kinases or functional kinase domains, as presently claimed in the

'781 application. It is also my opinion that the successful arraying of the active human

protein kinases by the Protometrix researchers would have reasonably been expected by

a scientist of ordinary skill in proteomics reading the '781 application on May 4, 2001, in

view of the reliance of the Protometrix researchers on the teaching of the '781

application in making and using these arrays.

28.   In the second Office Action, at page 20, the Patent Office quotes the

following statement made in one of my publications:

> [t]he family of human protein kinases consists of more than 500 members
> of which only a fraction have been characterized to date. Much is still not
> known about the biological function of many kinases, the protein
> substrates that are phosphorylated by these kinases, or the roles of these
> kinases and substrates in disease....[38]

The Patent Office then comments:

> [t]hus, Schweitzer has not extrapolated or predicted its findings to any
> other family [of] human protein kinases which consist of more than 500
> members to which only a fraction has been characterized to date.

Although not clearly articulated, it appears that the Patent Office is asserting I did not

extrapolate or predict that the successful arraying of human protein kinases described in

my first Declaration could be reasonably extended to arraying other, poorly characterized

---

[38] My first Declaration, Exhibit C, at page 2, column 1, paragraph 2 to page 3,
column 1, paragraph 1 (2004) (Exhibit 6).

human protein kinases. In response, I first point out that Patent Office has relied on the

quoted excerpt out of context. Specifically, I note that the fact much may not be known

about the *biological function, substrates, and roles of protein kinase* and their substrates

in disease, does *not* support the position that protein kinase family members (and their

fragments, having a functional kinase domain) would not be expected to display *kinase*

activity when arrayed according to the teaching of the '781 application. Moreover, the

Patent Office's reliance on this argument disregards my previous testimony that the

Protometrix arrays contain well over 122 active human kinases and functional kinase

domains.[39] Thus, a high percentage of *all* the protein kinases in the human kinome are

arrayed as active kinases on the Protometrix arrays (i.e., despite the fact that "only a

fraction have been characterized to date"). To directly address the point raised by the

Patent Office, it is my opinion that a scientist of ordinary skill, having read the '781

application on May 4, 2001, would have reasonably expected that by applying the

disclosure and teaching of the '781 application, well characterized, poorly characterized,

and uncharacterized protein kinases could be routinely and successfully arrayed as active

kinases. Moreover, on May 4, 2001, a scientist of ordinary skill in the field of

proteomics would indeed have reasonably expected that by applying the disclosure and

teaching of the '781 application, at least 61 protein kinases could similarly be routinely

and successfully arrayed and thus, would likewise display the requisite protein kinase

---

[39] My first Declaration, at Paragraph 13 (Exhibit 6). As an addendum to and in clarification of, my previous testimony, I note that while to the best of my knowledge Protometrix researchers have not formally rigorously proven that all 400 of the human kinases and functional kinase domains arrayed on Invitrogen's Human ProtoArray High Density Protein Microarrays™ are active, it is my belief and understanding based on autophosphorylation studies that a high percentage of these arrayed proteins are active (i.e., well over the 122 purified active human kinases or human functional kinase domains recited in pending claim 186 (Exhibit 3)).

activity, irrespective of whether the kinases are: (a) known or yet to be discovered; (b) well characterized, poorly characterized or uncharacterized; (c) full-length or a fragment containing a kinase domain, or (d) from a yeast, a *Drosophila*, or a human or another mammal.

Summary

29.   The '781 application contains extensive disclosure and guidance relating to the manufacture and use of arrays that fall within the scope of the pending claims.[40] In particular, Example 1 discloses methods and techniques that successfully produced 17 arrays containing 111 purified distinct and diverse active protein kinases corresponding to almost every protein kinase in the yeast kinome, of which a high percentage of the arrayed proteins displayed kinase activity.

30.   For at least the reasons presented in this declaration, it is my opinion that in view of the teaching and guidance of the '781 application and the extensive knowledge in the field of proteomics on May 4, 2001, which included an art-recognized correlation between the sequence (i.e., structure) and kinase activity (i.e., function) of kinase domains, a scientist of ordinary skill in the field of proteomics, enlightened by the disclosure of the '781 application would have reasonably concluded that:

(a)   protein kinases from different organisms were readily and reliably recognized and distinguished based on the primary amino acid sequences of their catalytic kinase domain and proteins containing these kinase domains were expected to display kinase activity;

---

[40] Pending claims (Exhibit 3).

(b)   characterized and uncharacterized or unidentified protein kinases (and

fragments containing a functional kinase domain of these proteins) from a yeast or other

organism, such as a *Drosophila*, or a mammal (including a human) could be readily

recognized, distinguished, and arrayed in accordance with the disclosure and teaching of

the '781 application; and

(c)   the '781 application contains extensive disclosure and guidance that

provides ample description of methods for reliably and predictably making and using

protein arrays falling within the scope of claim 1 and that contain at least 61 purified

active protein kinases (including fragments containing a functional kinase domain) from

a yeast or another organism, such as a *Drosophila* or a mammal (including a human), and

the arrays containing these purified active kinases.

31.   Accordingly, for at least the above reasons, it is my opinion that a

scientist of ordinary skill in the field of proteomics on May 4, 2001, reading the '781

application, would have reasonably concluded that the application amply describes the

claimed arrays which contain at least 61 purified active kinases (and fragments

containing a functional kinase domain) of a yeast, a *Drosophila* or a mammal (in

mammal).

32.   Additionally, for at least the reasons presented herein, it is my opinion

that in view of the extensive knowledge in the field of proteomics and the art-recognized

correlation between the sequence and kinase activity of kinase domains on May 4, 2001,

a scientist of ordinary skill in the field of proteomics, enlightened by the teaching

disclosure and guidance of the '781 application would have reasonably concluded that:

(a)   protein kinases (and fragments containing a functional kinase domain)

from different organisms were and could be routinely, and predictably recognized and

distinguished from other proteins based on the sequences of their catalytic kinase domain and proteins containing these kinase domains would be expected to display kinase activity that could readily be assayed using methods disclosed in the '781 application or otherwise known in the art; and

(b)   the extensive disclosure, teaching, and guidance of the '781 application, which includes methods and reagents used to successfully array purified active protein kinases corresponding to more than 90% of the yeast kinome, could be applied to routinely and predictably produce arrays encompassed by the pending claims and that in particular, contain at least 61 purified active protein kinases (including fragments containing a functional kinase domain) from a yeast and other organisms, such as a *Drosophila* or a mammal (including a human).

33.   For at least the above reasons, a scientist of ordinary skill in the art of proteomics, enlightened by the extensive guidance and teaching of the '781 application, would have reasonably concluded that the methods and techniques disclosed in the application could be routinely applied (as indeed they were) to make and use equally successful arrays falling within the scope of the pending claims (Exhibit 3) and that contain at least 61 purified active protein kinases or fragments containing function kinase domains, from a yeast, a *Drosophila,* or a mammal (including a human), without having to undertake more than routine experimentation.

34.   I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements

and the like so made are punishable by fine or imprisonment, or both, under Section
1001 of Title 18 of the United States Code and that such willful false statements may
jeopardize the validity of the present patent application or any patent issued thereon.

Respectfully submitted,

Barry Schweitzer, Ph.D.

Date: 4/26/11

1353227_2.DOC

Exhibit 1

BARRY SCHWEITZER, Ph.D.

CONTACT INFORMATION

459 Maple Avenue
Cheshire, CT 06410
203-439-9439 (Work)
203-645-2953 (Mobile)
barry.schweitzer@lifetech.com

PROFESSIONAL EXPERIENCE

INVITROGEN CORPORATION (Now LIFE TECHNOLOGIES), Carlsbad, CA   2004 - Present
Director, Integrated Technologies, Molecular Biology Systems Division – 2009 to present
Director, Protein Analysis R&D - 2008
Director, Protein Array R&D and Site Leader - 2006 – 2007
Director, Protein Array R&D and Operations - 2004 - 2006
Current responsibilities include the oversight of programs which span the traditional segments of the Molecular Biology Reagent Business, particularly programs that integrate instrumentation with consumables. Previous responsibilities included oversight of R&D and Services for Invitrogen's Protein Analysis product lines, including protein separation technologies, Western technologies, mass spectroscopy, and protein arrays. Additional responsibilities included site leadership of the Protein Array Center in Branford, CT, including R&D, Services, Manufacturing, Quality, and Facilities functions. Other responsibilities include budget preparation and implementation, intellectual property management, and oversight of academic, government, and industrial collaborations and contracts. Also participating in technology and intellectual property evaluations, business development, grant preparations, community relations, and presentations at national and international meetings. Reporting to the Vice President, R&D of the Molecular Biology Reagents Business Unit.

Leadership accomplishments include:
- Led transfer of all operations from Branford, CT to Carlsbad on-time, under budget, and without loss of revenue
- Led global launch of several new multimillion dollar products
- Championed Lean Six Sigma Black Belt and Green Belt projects
- Led ISO 9001 Certification of Branford Site
- Led successful completion of multimillion dollar Biodefense projects in partnership with the United States Army Medical Research Institute for Infectious Diseases (USAMRIID)
- Authored or co-authored 11 publications, including paper in *Nature*
- Inventor or co-inventor on 10 new patent applications
- Presented at 14 international scientific conferences.

PROTOMETRIX, INC., Branford, CT   2002 - 2004
Senior Director, Technology - 2003-2004
Director of Technology – 2002 – 2003
Fifth person to join start-up biotechnology company. Director of a research and development operation providing high-throughput gene cloning, protein expression, protein purification, and protein microarray manufacturing for products, services, and discovery. Additional responsibilities included leading product development teams, leading technology and intellectual property diligence reviews, presenting to investors, coordinating industrial collaborations, and managing prosecution of company intellectual property. Reported to the Vice President, R&D.

Leadership accomplishments included:
- Led the Protometrix technical and IP diligence team during the acquisition of the company by Invitrogen Corp.
- Led the commercial launch of the world's first functional protein microarray product.
- Established the 1st manufacturing facility for the production of protein arrays.
- Built highly skilled team of scientists, engineers, and informatics specialists
- Led the design and buildout of 14,000 s.f. state-of-the-art laboratory and company headquarters.

MOLECULAR STAGING, INC., New Haven, CT   1998 - 2002
Director of Proteomics - 2001 – 2002
Section Head - 1998-2000
Second person to join start-up biotechnology company. Director of a research and service operation providing high-throughput protein expression profiling data using proprietary protein microarray technology to academic, government, and corporate clients. Responsibilities included management of research personnel, budget preparation and implementation, business development, oversight of academic collaborators, preparation of publications and patent applications, presentations for investors, corporate partners and at national meetings. Four direct and 23 indirect reports. Reporting to Chief Operating Officer.

Leadership accomplishments included:
- Successfully launched the world's first microarray-based protein expression profiling service.
- Developed the world's most advanced manufacturing facility for production of antibody microarrays.
- 8 publications, including publication in Nature Biotechnology of 1st application of antibody microarrays for protein expression profiling.
- 1 issued patent, and 4 patent applications.
- Led and coordinated the design and buildout of 46,000 s.f. of state-of-the-art proteomics laboratory.
- Successfully moved an academic technology into an industrial setting, increasing sensitivity, robustness, and utility.
- Led project resulting in $9 MM equity investment by Fortune 100 Company.
- Gave technical presentations resulting in $40 MM 2nd round financing.

WALT DISNEY MEMORIAL CANCER CENTER, Orlando, FL   1994 - 1998
Division Director. Laboratory director of multidisciplinary research program in the structural biology of nucleic acids, proteins, and drugs involved in cancer and related diseases. Responsibilities included carrying out experiments and data analysis, project development, management of 15-20 research, administrative, and volunteer personnel, budget preparation and implementation, grant writing, preparation of publications, public relations, and mentoring of graduate, undergraduate and high school students.
Scientific Director Molecular Diagnostics Clinical Laboratory. Responsibilities included business plan preparation and implementation, management of technical staff, technical consultant, clinical research director, and physician outreach.

Leadership accomplishments included:
- Established and directed a program utilizing multidimensional nuclear magnetic resonance (NMR) spectroscopy, and computational chemistry to determine high-resolution structures of proteins, nucleic acids, and drug complexes for the purpose of chemotherapeutic development.
- Established and directed a laboratory utilizing the most advanced molecular techniques to diagnose infectious diseases, cancer, and inherited diseases for patients of Florida Hospital (2nd largest number of admissions in U.S.).

UNIVERSITY OF CENTRAL FLORIDA, Orlando, FL   1994 - 1998
Assistant Professor
Responsibilities included: Research, Florida Hospital liaison, committee service, mentoring of graduate and undergraduate students, taught courses in Principals of Modern NMR Spectroscopy, Special Topics in Drug Development, Advanced Biochemistry Laboratory

EARLIER POSITIONS: Associate Research Scientist (1991-1993), Yale University School of Medicine, and **Research Associate** (1990-1991), Memorial Sloan-Kettering Cancer Center

---

OTHER EXPERIENCE

GLYGENIX, INC., Cheshire, CT   2005 – 2007
Member, Board of Directors. Glygenix, Inc. was established to benefit children born with Glycogen Storage Disease, Type I (GSD1.) Its goal is to help find a cure for this disease by raising monies for GSD1-related research.

THE EPISCOPAL CHURCH AT YALE, New Haven, CT   2000 - 2003
Member, Board of Governors. The Episcopal Church at Yale (ECY) is a full time ministry of the Episcopal Church to students, staff and faculty at Yale.The ECY is governed by a Board of Governors of the Episcopal Church at Yale Corporation which is the legal entity of the Corporation in matters of contracts and other transactions with other institutions such as Yale University.

PUBLICATIONS

Original Articles.

1. Schweitzer, B. I., and Bacopoulos, N. G. (1983) Reversible decrease in dopaminergic 3H-agonist binding after 6-hydroxydopamine and irreversible decrease after kainic acid. *Life Sciences* 32, 531-541.

2. Merker, M., Rice, J., Schweitzer, B., and Handschumacher, R. E. (1983) Cyclosporine binding component in BW5187 lymphoblasts and normal lymphoid tissue. *Transplantation Proceedings* 15, 2265-2270.

3. DeReimer, S. A., Schweitzer, B., and Kaczmarek, L. K. (1985) Inhibitors of calcium-dependent enzymes prevent the onset of afterdischarge in the peptidergic bag cell neurons of Aplysia. *Brain Research* 340, 175-180.

4. Srimatkandada, S., Schweitzer, B. I., Moroson, B. A., Dube, S., and Bertino, J. R. (1989) Amplification of a polymorphic dihydrofolate reductase gene expressing an enzyme with decreased binding to methotrexate in a human colon carcinoma cell line, HCT8R4, resistant to this drug. *Journal of Biological Chemistry* 264, 3524-3528.

5. Schweitzer, B. I., Srimatkandada, S., Gritsman, H., Sheridan, R., Venkataraghavan, R., and Bertino, J. R. (1989) Probing the role of two hydrophobic residues in the active site of the human dihydrofolate reductase by site-directed mutagenesis. *Journal of Biological Chemistry* 264, 20786-20795.

6. Dicker, A. P., Volkenandt, M., Schweitzer, B. I., Banerjee, D., and Bertino, J. R. (1990) Identification and characterization of a mutation in the dihydrofolate reductase gene from the methotrexate resistant Chinese hamster ovary cell line, Pro-3 MTXRIII. *Journal of Biological Chemistry* 265, 8317-8321.

7. Li, W. W., Lin, J. T., Schweitzer, B. I., and Bertino, J. R. (1991) Mechanisms of sensitivity and natural resistance to antifolates in a methylcholanthrene-induced rat sarcoma. Molecular Pharmacology 40, 854-858.

8. Li, W. W., Lin, J. T., Chang, Y. M., Schweitzer, B. I., and Bertino, J. R. (1991) Prediction of antifolate efficacy in a rat sarcoma model. *International Journal of Cancer* 49, 234-238.

9. Lin, J. T., Tong, W. P., Trippett, T. M., Niedzwiecki, D., Tao, Y., Tan, C., Steinherz, P., Schweitzer, B. I., and Bertino, J. R. (1991) Basis for natural resistance to methotrexate in human acute non-lymphocytic leukemia. *Leukemia Research* 15, 1191-1196.

10. Sapse, A.-M., Schweitzer, B. I., Dicker, A. P., Bertino, J. R., and Frecer, V. (1992) Ab initio studies of aromatic-aromatic and aromatic-polar interactions in the binding of substrate and inhibitor to dihydrofolate reductase. *International Journal of Peptide and Protein Research* 339, 18-23.

11. Trippett, T., Schlemmer, S., Elisseyeff, Y., Wachter, M., Steinherz, P., Berman, E., Rosowsky, A., Schweitzer, B., and Bertino, J. R. (1992) Evidence for defective transport as a mechanism of acquired resistance to MTX in patients with acute lymphocytic leukemia. *Blood* 80, 1158-1162.

12. Li, W. W., Lin, J. T., Schweitzer, B. I., Tong, W. P., Niedzwiecki, D., Brennan, M. F., and Bertino, J. R. (1992) Intrinsic resistance to methotrexate in human soft tissue sarcoma cell lines. *Cancer Research* 52, 1-6.

13. Li, M.-X., Hantzopoulos, P. A., Banerjee, D., Zhao, S. C., Schweitzer, B. I., Gilboa, E., and Bertino, J. R. (1992) Comparison of the expression of a mutant dihydrofolate reductase under control of different internal promoters in retroviral vectors. *Human Gene Therapy* 3, 381-390.

14. Volkenandt, M., Dicker, A. P., Banerjee, D., Fanin, R., Schweitzer, B. I., Horikoshi, T., Danenberg, K, Danenberg P., and Bertino, J. R. (1992) Quantitation of gene copy number and mRNA using the polymerase chain reaction. *Proceedings of the Society for Experimental Biology and Medicine* 200, 1-6.

15. Fanin, R., Banerjee, D., Volkenandt, M., Waltham, M., Li, W. W., Dicker, A. P., Schweitzer, B. I., and Bertino, J. R. (1993) Mutations leading to antifolate resistance in Chinese hamster ovary cells after exposure to the alkylating agent ethylmethanesulfonate. *Molecular Pharmacology* 44, 13-21.

16. Goker, E., Lin, J. T., Trippet, T., Elisseyeff, Y., Tong, W. P., Niedzwiecki, D., Tan, C., Steinherz, P., Schweitzer, B. I., and Bertino, J. R. (1993) Decreased polyglutamylation of methotrexate in acute lymphoblastic leukemia blasts in adults compared to children with this disease. *Leukemia* 7, 1000-1004.

17. Li, W. W., Waltham, M., Tong, W., Schweitzer, B. I., and Bertino, J. R. (1993) Increased activity of gamma-glutamyl hydrolase in human sarcoma cell lines: a novel mechanism of intrinsic resistance to methotrexate (MTX). *Advances in Experimental Medicine and Biology* 338, 635-638.

18. Kellogg, G. W., and Schweitzer, B. I. (1993) Two- and three-dimensional 31P-driven NMR procedures for complete assignment of backbone resonances in oligodeoxyribonucleotides. *Journal of Biomolecular NMR* 3, 577-595.

19. Dicker, A. P., Waltham, M., Volkenandt, M., Schweitzer, B. I., Otter, G. M., Schmid, F. A., Sirotnak, F. M., and

Bertino, J. R. (1993) Methotrexate resistance in an in vivo mouse tumor resulting from a novel mutation in the dihydrofolate reductase gene. *Proceedings of the National Academy of Sciences* (USA) 90, 11797-11801.

20. Ercikan, E., Waltham, M., Dicker, A., Schweitzer, B., and Bertino, J. R. (1993) Effect of codon 22 mutations on substrate and inhibitor binding for human dihydrofolate reductase. *Advances in Experimental Medicine and Biology* 338, 515-519.

21. Banerjee, D., Schweitzer, B. I., Volkenandt, M., Li, M.-X., Waltham, M., Mineishi, S., Zhao, S.-C., and Bertino, J. R. (1994) Transfection with a cDNA encoding a Ser31 or Ser34 mutant human dihydrofolate reductase into Chinese hamster ovary and mouse marrow progenitor cells confers methotrexate resistance. *Gene* 139, 269-274.

22. Li, M.-X., Banerjee, D., Zhao, S.-C., Schweitzer, B. I., Mineishi, S., Gilboa, E., and Bertino, J. R. (1994) Development of a retroviral construct containing a human mutated dihydrofolate reductase cDNA for hematopoietic stem cell transduction. *Blood* 83, 3403-3408.

23. Zhao, S.-C., Li, M.-X., Banerjee, D., Schweitzer, B. I., Mineishi, S., Gilboa, E., and Bertino, J. R. (1994) Long-term protection of recipient mice from lethal doses of methotrexate by marrow infected with a double-copy vector retrovirus containing a mutant dihydrofolate reductase. *Cancer Gene Therapy* 1, 33-39.

24. Schweitzer, B. I., Mikita, T., Kellogg, G. W., Gardner, K. H., and Beardsley, G. P. (1994) Solution structure of a DNA dodecamer containing the antineoplastic agent cytosine arabinoside: Determination by two- and three-dimensional NMR, restrained molecular dynamics, and complete relaxation matrix analysis. *Biochemistry* 33, 11460-11475.

25. Schweitzer, B. I., Gardner, K. H., and Tucker-Kellogg, G. (1995) HeteroTOCSY-based experiments for measuring heteronuclear relaxation in nucleic acids and proteins. *Journal of Biomolecular NMR* 6, 180-188.

26. Marshalko, S. J., Schweitzer, B. I., and Beardsley, G. P. (1995) Chiral chemical synthesis of DNA containing (S)-9-(1,3-dihydroxy-2-propoxymethyl) guanine (DHPG) and effects on thermal stability, duplex structure, and thermodynamics of duplex formation. *Biochemistry* 34, 9235-9248.

27. Callihan, D., West, J., Schweitzer, B. I., and Logan, T. M. (1996) Simple, distortion-free homonuclear spectra of peptides and nucleic acids in water using excitation sculpting. *Journal of Magnetic Resonance*, Series B, 112, 82-86.

28. Erickan, A. E., Waltham, M., Dicker, A. P., Schweitzer, B. I., Gritsman, H., Banerjee, D,. and Bertino, J. R. (1996) Variants of human dihydrofolate reductase with substitutions at leucine-22: effect on catalytic and inhibitor binding properties. *Molecular Pharmacology* 49, 430-437.

29. Foti, M., Marshalko, S. J., Beardsley, G. P., and Schweitzer, B. I. (1997) Solution structure of a DNA decamer containing the antiviral agent ganciclovir: Determination by two-dimensional homonuclear and heteronuclear NMR, restrained molecular dynamics, and complete relaxation matrix analysis. *Biochemistry* 36, 5336-5345.

30. Kumar, S., Reed, M., Gamper, H., Gorn, V., Lukhtanov, E., Kutyavin, I., Meyer, R., and Schweitzer, B. I. (1997) Solution structure of a DNA decamer conjugated on the 5' end with CDPI3. *Nucleic Acids Research* 26, 831-838.

31. Shao, W., Jerva, L. F., West, J., Lolis, E., and Schweitzer, B. I. (1998) Solution structure of murine macrophage inflammatory protein-2 (MIP-2). *Biochemistry* 37, 8303-8313.

32. Shao, W., Fernandez, E., Wilken, J., Thompson, D. A., Siani, M. A., West, J., Lolis, E., and Schweitzer, B. I. (1998) Accessibility of selenomethionine proteins by total chemical synthesis: structural studies of human herpes virus-8 MIP-II. *FEBS Letters* 441, 77-82.

33. Schweitzer, B. I., Foti, M., Keertikar, K., Kumar, S., Gardner, K. H., and Tucker-Kellogg, G. (1999) The use of ^{31}P relaxation experiments to probe the effects of nucleoside analogs on DNA dynamics. *Phosphorus, Sulfur, and Silicon* 12, 36-41.

34. Foti, M., Omichinski, J. G., West, J., Stahl, S., and Schweitzer, B. I. (1999) Effects of nucleoside analog incorporation on DNA interactions with the DNA binding domain of the GATA-1 erythroid transcription factor. *FEBS Letters* 444, 47-53.

35. Schweitzer, B., Wiltshire, S., Lambert, J., O'Malley, S., Kukanskis, K., Zhu, Z., Ward, D., and Kingsmore, S. F. (2000) Immunoassays with rolling circle DNA amplification: a versatile platform for ultrasensitive antigen detection. *Proceedings of the National Academy of Sciences (USA)* 97, 10113-10119.

36. Wiltshire, S., O'Malley, S., Lambert, J., Kukanskis, K., Edgar, D., Kingsmore, S. F., and Schweitzer, B.I. (2000) Detection of multiple allergen-specific IgE on microarrays by immunoassay with rolling circle amplification. *Clinical Chemistry* 46, 1990-1993.

37. Shao, W., Fernandez, E., Wilken, J., Thompson, D. A., Schweitzer, B. I., and Lolis, E. (2001) CCR2 and CCR5 receptor-binding properties of herpesvirus-8 vMIP-II based on sequence analysis and its solution structure. *European Journal of Biochemistry* 268, 2948-2959.

38. Gusev, Y, Sparkowski, J., Ferguson, H., Montano, J., Visconti, R., Bogdan, N., Schweitzer, B., Wiltshire, S., Kingsmore, S., Maltzman, W., and Wheeler, V. (2001) Rolling circle amplification – A new approach to increase sensitivity for immunohistochemistry and flow cytometry. *Journal of American Pathology* 159, 63-69.

39. Mullenix, M. C., Wiltshire, S., Shao, W., Kitos, G., and Schweitzer, B. (2001) Allergen-specific IgE detection on microarrays using rolling circle amplification: correlation with in vitro assays for serum IgE. *Clinical Chemistry* 47, 1926-1929.

40. Nallur, G., Luo, C., Fang, L., Dave, V., Lambert, J., Lasken, R., Kingsmore, S., and Schweitzer, B. (2001) Signal amplification by rolling circle amplification on DNA microarrays. *Nucleic Acid Research* 29, e118.

41. Schweitzer, B., Roberts, S., Grimwade, B., Shao, W., Wang, M., Fu, Q., Shu, Q., Laroche, I., Christiansen, J., Velleca, M. and Kingsmore, S. F. (2002) Multiplexed protein profiling on microarrays by rolling circle amplification: Application to dendritic cell cytokine secretion. *Nature Biotechnology* 20, 259-365.

42. Nallur, G., Marrero, R., Luo, C., Krishna, R., Bechtel, P., Shao, W., Ray, M., Wiltshire, S, Fang, L, Huang, H., Liu, C.-G., Sun, L., Sawyer, J., Kingsmore, S., Schweitzer, B., and Xia, J. (2003) Protein and nucleic acid detection by rolling circle amplification on gel-based microarrays. *Biomedical Microdevices* 5, 115-123

43. Michaud, G, Salcius, M., Zhou, F., Bangham, R., Bonin, J., Guo, H., Snyder, M., Predki, P., and Schweitzer, B. (2003) Analyzing antibody specificity with whole proteome microarrays. *Nature Biotechnology* 21, 1509-1512.

44. De Yang, Q., Rosenberg, H., Rybak, S., Newton, D., Wang, Z., Fu, Q., Tchernev, V., Wang, M., Schweitzer, B., Kingsmore, S., Patel, D., Oppenheim, J., and Howard. O (2004) Human ribonuclease A superfamily members, eosinophil-derived neurotoxin (EDN) and pancreatic ribonuclease (hPR), induce dendritic cell maturation and activation. *Journal of Immunology* 173, 6134-6142.

45. Predki, P.F., Mattoon, D., Bangham, R., Schweitzer, B., and Michaud, G. Protein microarrays: A new tool for profiling antibody cross-reactivity. *Human Antibodies*. 2005, 14:7-15.

46. Ptacek, J., Devgan, G., Michaud, G., Zhu, H., Zhu, X., Fasolo, J., Guo, H., Jona, G., Breitkreutz, A., Sopko, R., McCartney, R., Schmidt, M., Rachidi, N., Lee, S.-J., Mah, A., Meng, L., Stark, M., Stern, D., De Virgilio, C., Tyers, M., Andrews, B., Gerstein, M., Schweitzer, B., Predki, P., and Snyder., M. (2005) Global analysis of protein phosphorylation in yeast. *Nature* 438, 679-84.

47. Boyle, S.N., Michaud, G.A., Schweitzer B., Predki, P.F., and Koleske, A.J. (2007) A critical role for cortactin phosphorylation by Abl-family kinases in PDGF-induced dorsal-wave formation. *Current Biology* 17, 445-51.

48. Meng, L., Michaud, G.A., Merkel, J.S., Zhou. F., Huang. J., Mattoon, D.R., and Schweitzer, B. (2008) Protein kinase substrate identification on functional protein arrays. *BMC Biotechnology*. 8, 22-30.

49. Schmid, K., Keasey, S., Pittman, P., Emerson, G.L., Meegan, J., Tikhonov, A.P., Chen, G., B. Schweitzer, B., and Ulrich, R.G. (2008) Analysis of the human immune response to vaccinia by use of a novel protein microarray suggests that antibodies recognize less than 10% of the total viral proteome. *Proteomics Clinical Applications* 2, 1528-1538.

50. Sboner, A., Karpikov, A., Chen, G., Smith, M., Mattoon, D., Freeman-Cook, L., Schweitzer, B., Gerstein, M.B. (2009) Robust linear model normalization to reduce technical variability in functional protein microarrays. *Journal of Proteome Research* 8, 5451-5464.

51. Keasey, S.L., Schmid, K.E., Lee, M.S, Meegan, J., Tomas, P., Minto, M., Tikhonov, A.P., Schweitzer, B., Ulrich, R.G. (2009) Extensive antibody cross-reactivity among infectious gram-negative bacteria revealed by proteome microarray analysis. *Molecular and Cellular Proteomics* 8, 924-935.

Reviews.

1. Bertino, J. R., Romanini, A., Dicker, A., Volkenandt, M., Lin, J. T., and Schweitzer, B. I. (1990) Resistance to methotrexate in experimental models and patients. In: Chemistry and Biology of Pteridines 1989, H. C. Curtius, S. Gisal, and N. Blau, Eds. New York: W. deGruyter, 1089-1100.

2. Schweitzer, B. I., Dicker, A. P., and Bertino, J. R., (1990) Dihydrofolate reductase as a therapeutic target. *FASEB Journal* 4, 2441-2452.

3. Bertino, J. R., Romanini, A., Dicker, A., Volkenandt, M., Lin, J. T., and Schweitzer, B. I. (1990) Resistance to MTX: experimental and clinical. In: Drug Resistance: Mechanisms and Reversal. E. Mihich, Ed. New York: John Libbey, 33-46.

4. Bertino, J. R., Li, W. W., Lin, J., Trippett, T., Goker, E., Schweitzer, B., and Banerjee, D. (1992) Enzymes of the

thymidylate cycle as targets for chemotherapeutic agents: mechanisms of resistance. *Mount Sinai Journal of Medicine* 59, 391-395.

5. Banerjee, D., Zhao. S. C., Li, M.-X., Schweitzer, B. I., Mineishi, S., and Bertino, J. R. (1994) Gene therapy utilizing drug resistance genes: a review. Stem Cells 12, 378-385.

6. Schweitzer, B., and Kingsmore, S. (2001) Combining nucleic acid amplification and detection. *Current Opinion in Biotechnology* 12, 21-27.

7. Schweitzer, B., and Kingsmore, S. (2002) Measuring proteins on microarrays. *Current Opinion in Biotechnology* 13, 14-19.

8. Schweitzer, B., Predki, P. and Snyder, M. (2003) Microarrays to characterize protein interactions on a whole-proteome scale. *Proteomics* 3, 2190-2199.

9. Zhou FX, Bonin J, and Predki PF (2004) Development of functional protein microarrays for drug discovery: progress and challenges. *Comb Chem High Throughput Screen* 7:539-46.

10. Merkel JS, Michaud GA, Salcius M, Schweitzer B, and Predki PF. (2005) Functional protein microarrays: just how functional are they? *Curr Opin Biotechnol.* 16:447-52.

11. Bangham R, Michaud GA, Schweitzer B, Predki PF. (2005) Protein microarray-based screening of antibody specificity. *Methods Mol Med.* 114:173-82.

12. Mattoon D, Michaud G, Merkel J, Schweitzer B. (2005) Biomarker discovery using protein microarray technology platforms: antibody-antigen complex profiling. *Expert Reviews in Proteomics* 2:879-89.

13. Michaud, G., Samuels, M., and Schweitzer B. (2006) Functional protein arrays to facilitate drug discovery and development. *IDrugs.* 9:266-72.

14. Michaud, G.A., Salcius, M., Zhou, F., Papov, V.V., Merkel, J., Murtha, M., Predki, P., and Schweitzer, B. (2006) Applications of protein arrays for small molecule drug discovery and characterization. *Biotechnol Genet Eng Rev.* 22:197-211.

15. Salcius, M., Michaud, G.A., Schweitzer, B., and Predki, P.F. (2007) Identification of small molecule targets on functional protein microarrays. *Methods Mol Biol.* 382:239-48.

16. Mattoon, D.R., and Schweitzer, B. (2009) Antibody specificity profiling on functional protein microarrays. *Methods Bol. Biol.* 524: 213-23

Exhibit 2

Please type a plus (+) inside this box

05-07-01

UTILITY PATENT APPLICATION TRANSMITTAL

(Only for new nonprovisional applications under 37 CFR 1.53(b))

Attorney Docket No.	6523-028	Total Pages	71
First Named Inventor or Application Identifier			
Michael Snyder			
Express Mail Label No.	EL 501 638 671 US		

ADDRESS TO: Assistant Commissioner for Patents
Box Patent Application
Washington, DC 20231

APPLICATION ELEMENTS
See MPEP chapter 600 concerning utility patent application contents.

1. ☒ Fee Transmittal Form
   Submit an original, and a duplicate for fee processing)

2. ☒ Specification   [*Total Pages* __59__]
   (preferred arrangement set forth below)
   -Descriptive title of the Invention
   -Cross Reference to Related Applications
   -Statement Regarding Fed sponsored R&D
   -Reference to Microfiche Appendix
   -Background of the Invention
   -Brief Summary of the Invention
   -Brief Description of the Drawings *(if filed)*
   -Detailed Description of the Invention (including drawings, *if filed*)
   -Claim(s)
   -Abstract of the Disclosure

3. ☒ Drawing(s) *(35 USC 113)*   [*Total Sheets* __8__]

4. ☒ Oath or Declaration   [*Total Sheets* __2__]
   a. ☒ Unexecuted (original or copy)
   b. ☐ Copy from a prior application (37 CFR 1.63(d))
      (for continuation/divisional with Box 17 completed)
      [Note Box 5 below]
      i. ☐ DELETION OF INVENTORS(S)
         Signed statement attached deleting inventor(s) named in the prior application, see 37 CFR 1.63(d)(2) and 1.33 (b).

5. ☐ Incorporation By Reference *(useable if Box 4b is checked)*
   The entire disclosure of the prior application, from which a copy of the oath or declaration is supplied under Box 4b, is considered as being part of the disclosure of the accompanying application and is hereby incorporated by reference therein.

6. ☐ Microfiche Computer Program *(Appendix)*

7. ☐ Nucleotide and/or Amino Acid Sequence Submission
   (if applicable, all necessary)
   a. ☐ Computer Readable Copy
   b. ☐ Paper Copy (identical to computer copy)
   c. ☐ Statement verifying identity of above copies

ACCOMPANYING APPLICATION PARTS

8. ☐ Assignment Papers (cover sheet & document(s))
9. ☐ 37 CFR 3.73(b) Statement   ☐ Power of Attorney *(when there is an assignee)*
10. ☐ English Translation Document *(if applicable)*
11. ☐ Information Disclosure   ☐ Copies of IDS Statement (IDS)/PTO-1449   Citations
12. ☐ Preliminary Amendment
13. ☒ Return Receipt Postcard (MPEP 503) *(Should be specifically itemized)*
14. ☐ Small Entity   ☐ Statement filed in prior application, Statement(s) Status still proper and desired
15. ☐ Certified Copy of Priority Document(s) *(if foreign priority is claimed)*
16. ☐ Other:

17. If a **CONTINUING APPLICATION**, *check appropriate box and supply the requisite information:*
   ☐ Continuation   ☐ Divisional   ☐ Continuation-in-part (CIP)   of prior application No: filed .

18. CORRESPONDENCE ADDRESS

☒ Customer Number or Bar Code Label
(Insert Customer No. or Attach bar code label here)   or ☐ Correspondence address below

NAME	
ADDRESS	

CITY		STATE		ZIP CODE	
COUNTRY		TELEPHONE		FAX	

NY2 - 1195861.1

PENNIE & EDMONDS LLP

COUNSELLORS AT LAW

1155 Avenue of the Americas
New York, N.Y. 10036-2711
(212) 790-9090

ATTORNEY DOCKET NO. __6523-028__   Date: __May 4, 2001__

Assistant Commissioner for Patents
Box PATENT APPLICATION
Washington, D.C. 20231

Sir:

The following utility patent application is enclosed for filing:

Applicant(s):   Michael Snyder, Mark Reed, Hen Zhu, and James   Executed on: Unexecuted
                Frank Klemic

Title of Invention: PROTEIN CHIPS FOR HIGH THROUGHPOUT SCREENING OF PROTEIN ACTIVITY

PATENT APPLICATION FEE VALUE

TYPE	NO. FILED	LESS	EXTRA	EXTRA RATE	FEE	
Total Claims		-20	0	$18.00 each	$	0.00
Independent		-3	0	$80.00 each	$	0.00
			Minimum Fee		$	0.00
			Multiple Dependency Fee If Applicable ($270.00)		$	0.00
			Total		$	0.00
		50% Reduction for Independent Inventor, Nonprofit Organization or Small Business Concern (a verified statement as to the applicant's status is attached)			- $	0.00
			Total Filing Fee		$	0.00

☐ DO NOT PUBLISH. I hereby certify that the invention disclosed in the attached application has not and will not be the subject of an application filed in another country, or under a multilateral agreement, that requires publication at eighteen months after filing. I hereby request that the attached application not be published under 35 U.S.C. 122(b).

☒ Priority of application no. 60,201,921 filed on May 4, 2000 and 60/221,034 filed July 27, 2000 in the U.S. is claimed under 35 U.S.C. § 119(e).

☐ The certified copy of the priority application has been filed in application no. filed

☐ Amend the specification by inserting before the first line the following sentence: This is a continuation-in-part of application no. filed .

Respectfully submitted,

By _Adriane M. Antler_

Adriane M. Antler   32,605
PENNIE & EDMONDS LLP   (Reg. No.)

Enclosure

This form is not for use with continuation, divisional, re-issue, design or plant patent applications.
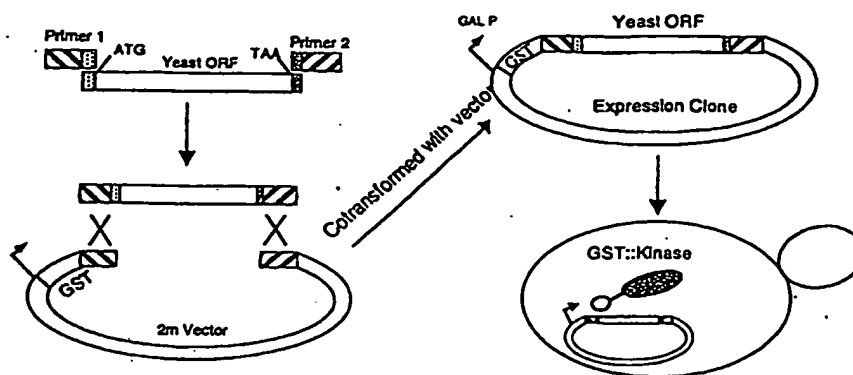
FIG. 1A

FIG. 1B

FIG. 2A

FIG. 2B

FIG. 3

6523 028

FIG.4A

FIG. 4B

FIG. 5A

Basic residue

DFG Sequence

Met

Asp

APE Sequence

Basic residue
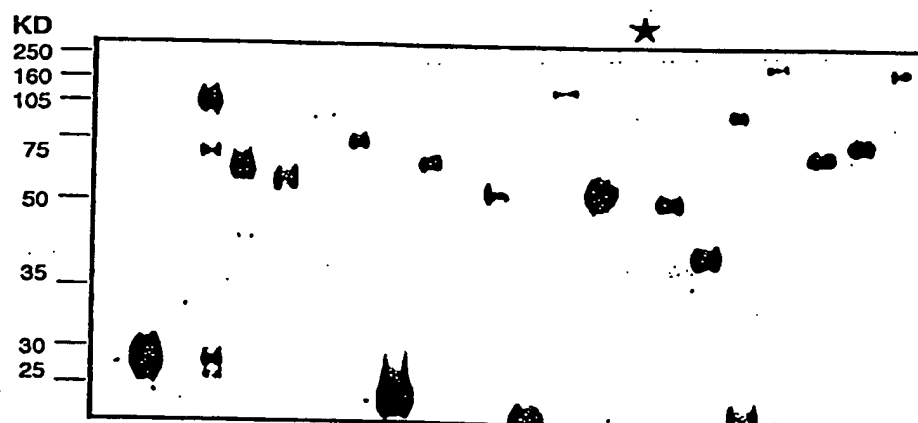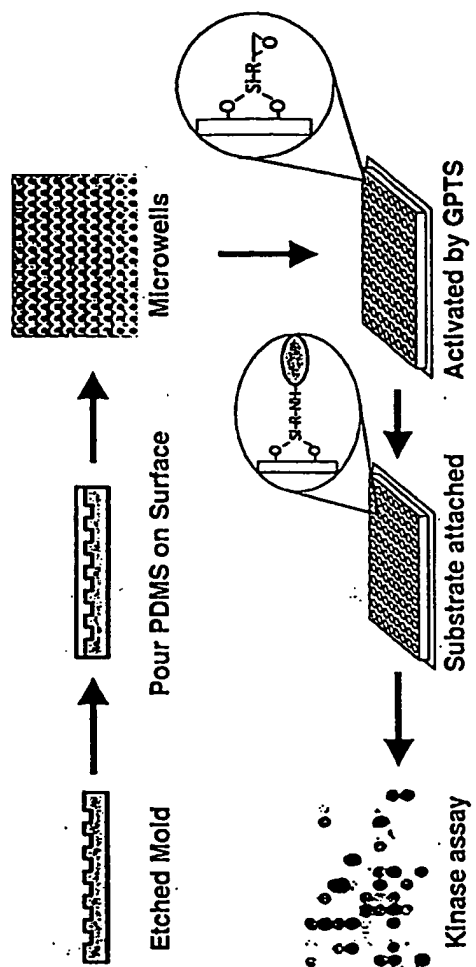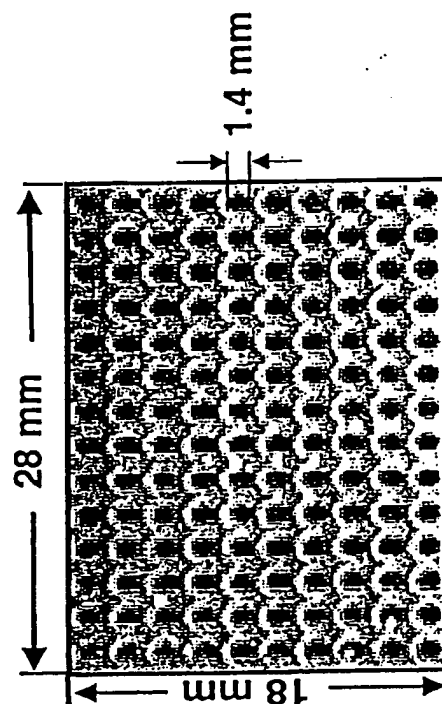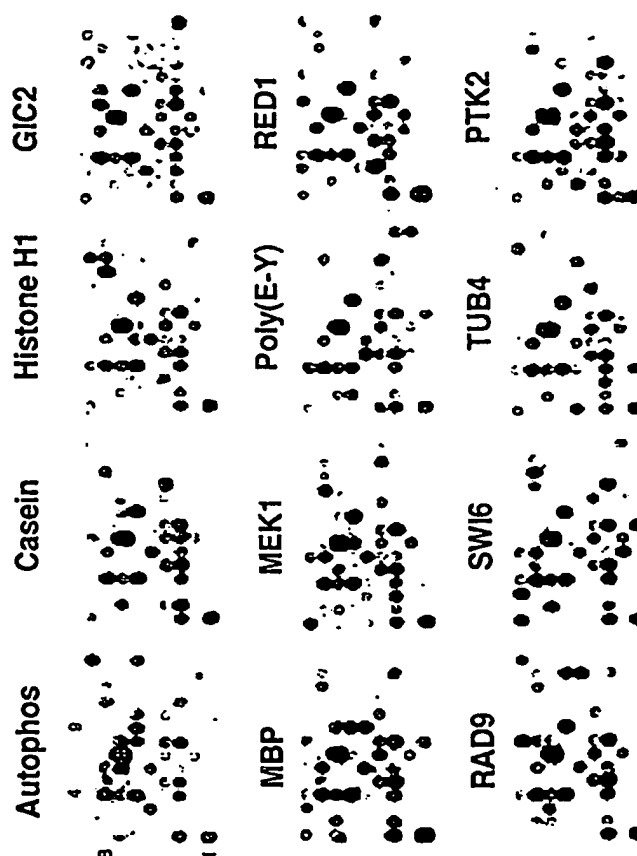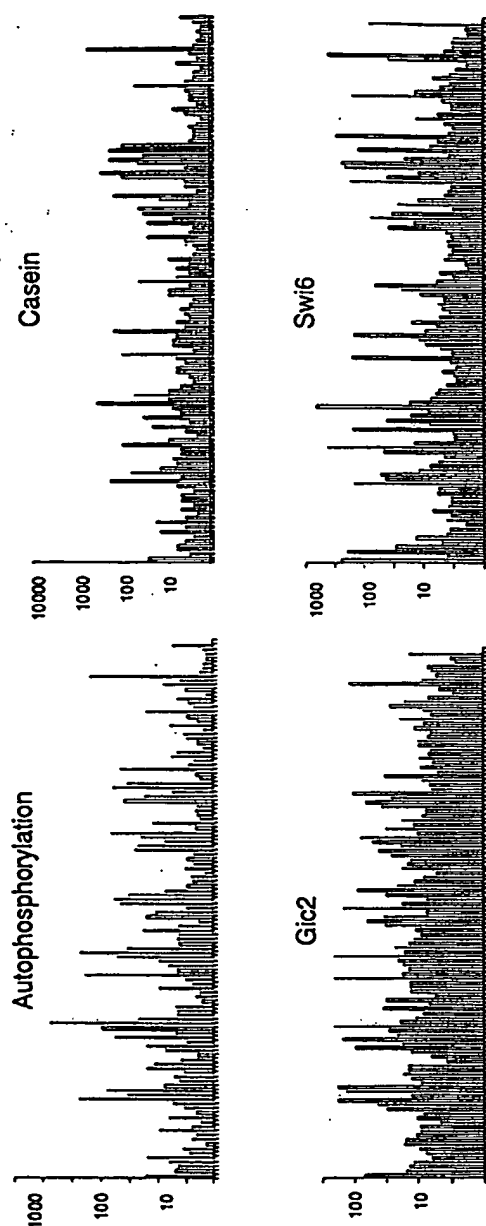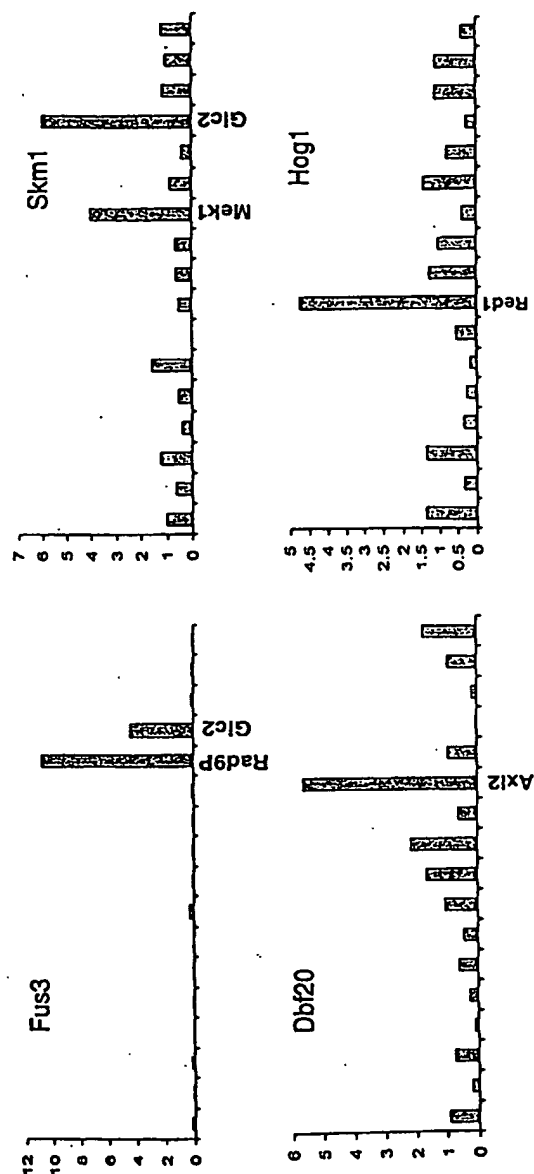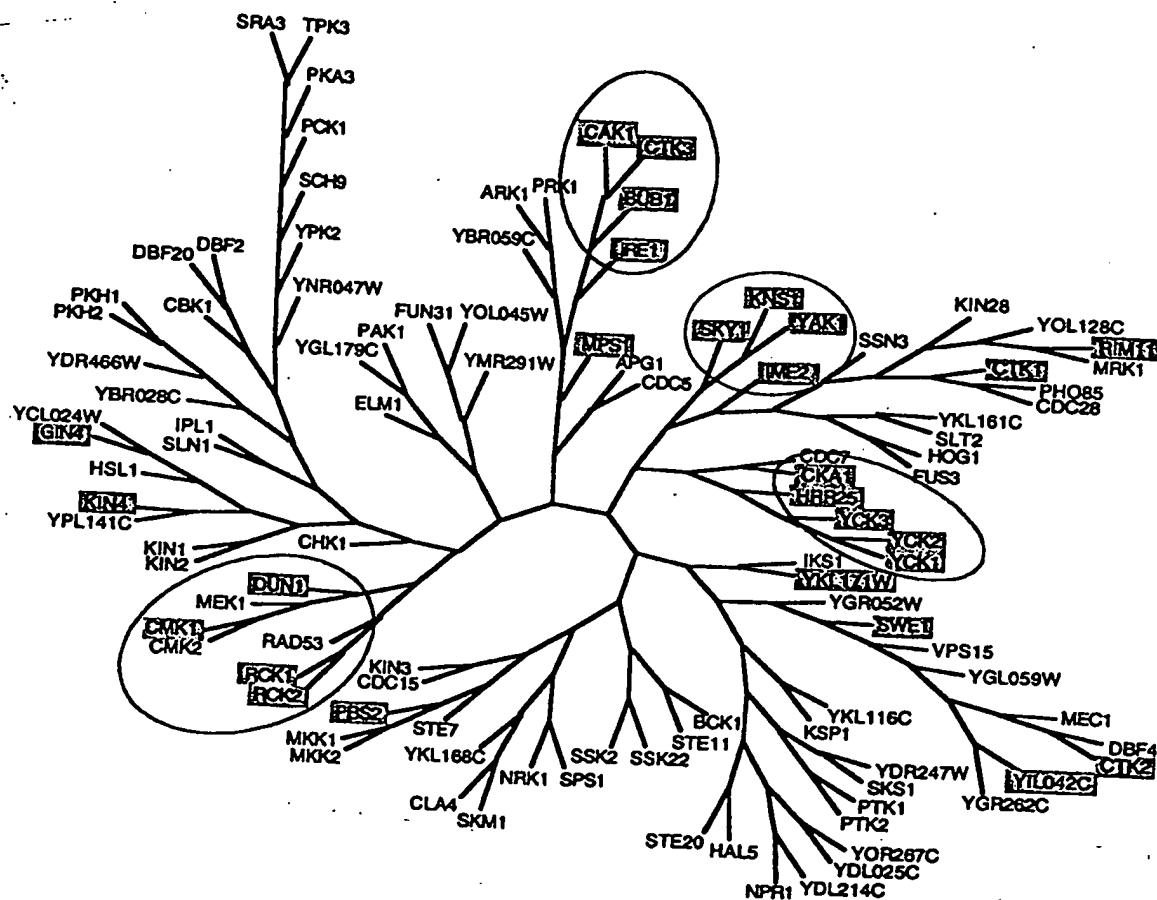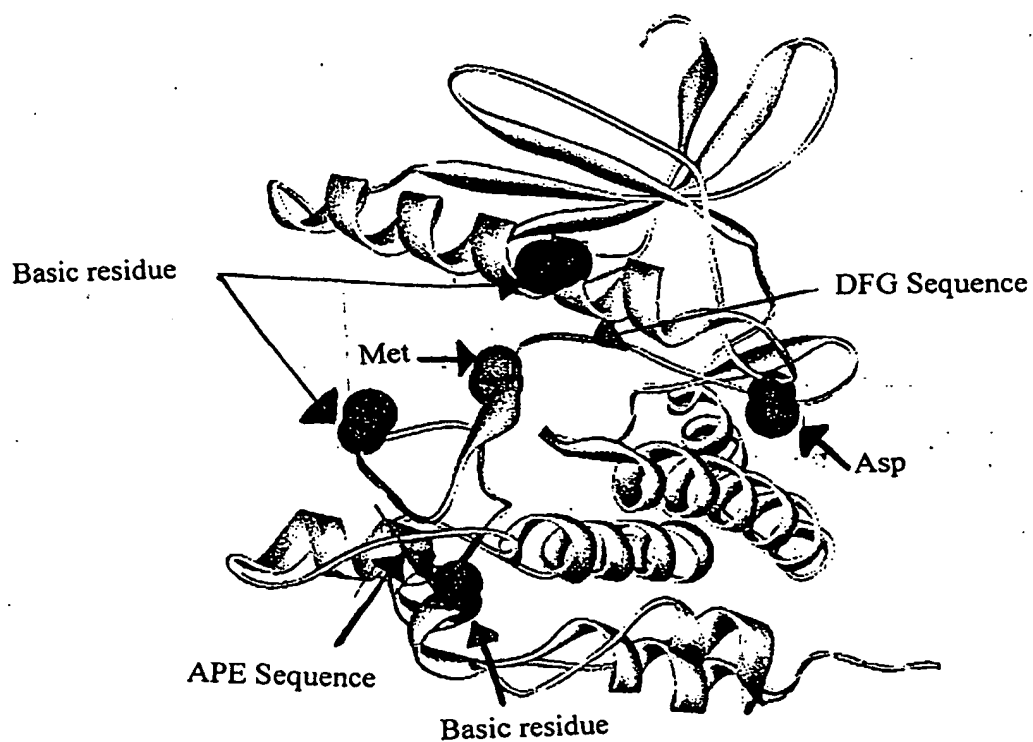
FIG.5B

6523 028

A

B.

C

D

E

F

FIG. 6

FIG. 7

5
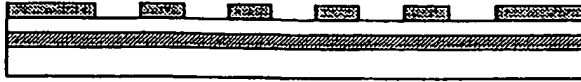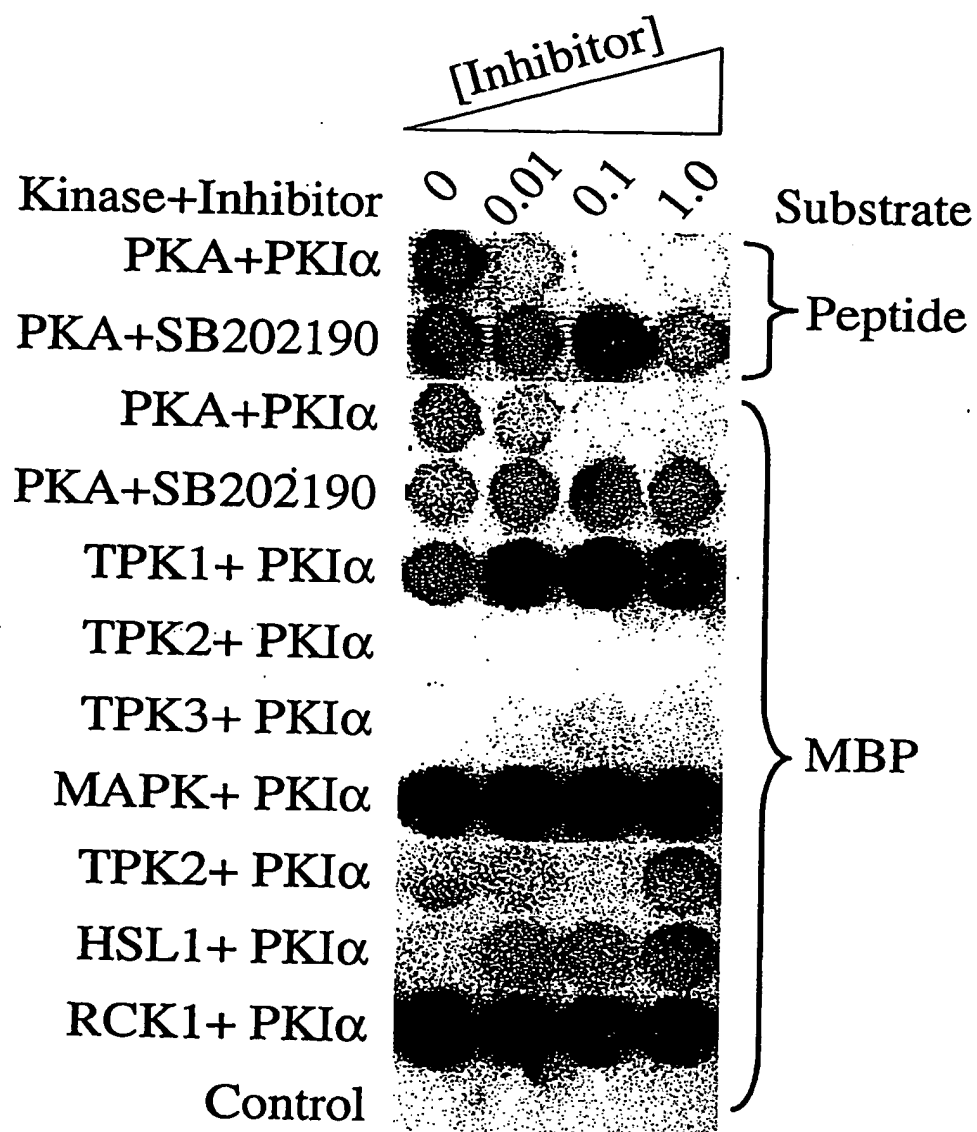
PROTEIN CHIPS FOR HIGH THROUGHPUT
SCREENING OF PROTEIN ACTIVITY

TABLE OF CONTENTS

5

10

15

20

25

30

35

PROTEIN CHIPS FOR HIGH THROUGHPUT
SCREENING OF PROTEIN ACTIVITY

5   This application claims the benefit under 35 U.S.C. § 119 (e) of U.S. provisional patent application Serial No. 60/201,921, filed on May 4, 2000, and U.S. provisional patent application Serial No. 60/221,034, filed on July 27, 2000, each of which is incorporated herein, by reference, in its entirety.

This invention was made with government support under grant numbers
10   DARPA/ONR R13164-41600099 and NIH (National Institutes of Health) RO1CA77808. The government has certain rights in the invention.

I.   Field of the Invention

15   The present invention relates to protein chips useful for the large-scale study of protein function where the chip contains densely packed reaction wells. The invention relates to methods of using protein chips to assay simultaneously the presence, amount, and/or function of proteins present in a protein sample or on one protein chip, or to assay the presence, relative specificity, and binding affinity of each probe in a mixture of probes
20   for each of the proteins on the chip. The invention also relates to methods of using the protein chips for high density and small volume chemical reactions. Also, the invention relates to polymers useful as protein chip substrates and methods of making protein chips. The invention further relates to compounds useful for the derivatization of protein chip substrates.

25

II.   Background of the Invention

The sequencing of entire genomes has resulted in the identification of large numbers of open reading frames (ORFs). Currently, significant effort is devoted to understanding
30   gene function by mRNA expression patterns and by gene disruption phenotypes. Important advances in this effort have been possible, in part, by the ability to analyze thousands of gene sequences in a single experiment using gene chip technology. However, much information about gene function comes from the analysis of the biochemical activities of the encoded protein.

35   Currently, these types of analyses are performed by individual investigators studying a single protein at a time. This is a very time-consuming process since it can take years to

purify and identify a protein based on its biochemical activity. The availability of an entire genome sequence makes it possible to perform biochemical assays on every protein encoded by the genome.

To this end, it would be useful to analyze hundreds or thousands of protein samples
5   using a single protein chip. Such approaches lend themselves well to high throughput experiments in which large amounts of data can be generated and analyzed. Microtiter plates containing 96 or 384 wells have been known in the field for many years. However, the size (at least 12.8 cm x 8.6 cm) of these plates makes them unsuitable for the large-scale analysis of proteins because the density of wells is not high enough.

10   As noted above, other types of arrays have been devised for use in DNA synthesis and hybridization reactions, e.g., as described in WO 89/10977. However, these arrays are unsuitable for protein analysis in discrete volumes because the arrays are constructed on flat surfaces which tend to become cross-contaminated between features.

Photolithographic techniques have been applied to making a variety of arrays, from
15   oligonucleotide arrays on flat surfaces (Pease et al., 1994, "Light-generated oligonucleotide arrays for rapid DNA sequence analysis," PNAS 91:5022-5026) to arrays of channels (U.S. Patent No. 5,843,767) to arrays of wells connected by channels (Cohen et al., 1999, "A microchip-based enzyme assay for protein kinase A," Anal Biochem. 273:89-97). Furthermore, microfabrication and microlithography techniques are well known in the
20   semiconductor fabrication area. See, e.g., Moreau, Semiconductor Lithography: Principals, Practices and Materials, Plenum Press, 1988.

Recently devised methods for expressing large numbers of proteins with potential utility for biochemical genomics in the budding yeast Saccharomyces cerevisiae have been developed. ORFs have been cloned into an expression vector that uses the GAL promoter
25   and fuses the protein to a polyhistidine (e.g., HISX6) label. This method has thus far been used to prepare and confirm expression of about 2000 yeast protein fusions (Heyman et al., 1999, "Genome-scale cloning and expression of individual open reading frames using topoisomerase I-mediated ligation," Genome Res. 9:383-392). Using a recombination strategy, about 85% of the yeast ORFs have been cloned in frame with a GST coding region
30   in a vector that contains the CUP1 promoter (inducible by copper), thus producing GST fusion proteins (Martzen et al., 1999, "A biochemical genomics approach for identifying genes by the activity of their products," Science 286:1153-1155). Martzen et al. used a pooling strategy to screen the collection of fusion proteins for several biochemical activities (e.g., phosphodiesterase and Appr-1-P-processing activities) and identified the relevant
35   genes encoding these activities. However, strategies to analyze large numbers of individual protein samples have not been described.

Thus, the need exists for a protein chip in which the wells are densely packed on the chip so as to gain cost and time advantage over the prior art chips and methods.

Citation or identification of any reference in Section II or any other section of this application shall not be considered as admission that such reference is available as prior art to the present invention.

III.   **Summary of the Invention**

The invention is directed to protein chips, *i.e.*, positionally addressable arrays of
10   proteins on a solid support, useful for the large-scale study of protein function wherein the protein chip contains densely packed reaction wells. The invention is also directed to methods of using protein chips to assay the presence, amount, and/or functionality of proteins present in at least one sample. The invention also is directed to methods of using the protein chips for high density and small volume chemical reactions. Also, the invention
15   is directed to polymers useful as protein chip substrates and methods of making protein chips. The invention is directed to compounds useful for the derivatization of protein chips.

In one embodiment, the present invention provides a protein chip comprising a flat surface, such as, but not limited to, glass slides. Dense protein arrays can be produced on, for example, glass slides, such that chemical reactions and assays can be conducted, thus
20   allowing large-scale parallel analysis of the presence, amount, and/or functionality of proteins. In a specific embodiment, the flat surface array has proteins bound to its surface via a 3-glycidooxypropyltrimethoxysilane (GPTS) linker.

Furthermore, in another specific embodiment, the present invention overcomes the disadvantages and limitations of the methods and apparatus known in the art by providing
25   protein chips with densely packed wells in which chemical reactions and assays can be conducted, thus allowing large-scale parallel analysis of the presence, amount, and/or functionality of proteins.

The general advantages of assaying arrays rather than one-by-one assays include the ability to simultaneously identify many protein-probe interactions, and to determine the
30   relative affinity of these interactions. The advantages of applying complex mixtures of probes to a chip include the ability to detect interactions in a milieu more representative of that in a cell, and the ability to simultaneously evaluate many potential ligands.

In one embodiment, the invention is a positionally addressable array comprising a plurality of different substances, selected from the group consisting of proteins, molecules
35   comprising functional domains of said proteins, whole cells, and protein-containing cellular material, on a solid support, with each different substance being at a different position on

- 3 -

the solid support, wherein the plurality of substances consists of at least 100 different substances per cm².

In another embodiment, the invention is a positionally addressable array comprising a plurality of different proteins, or molecules comprising functional domains of said proteins, on a solid support, with each different protein or molecule being at a different position on the solid support, wherein the plurality of different proteins or molecules consists of at least 50% of all expressed proteins with the same type of biological activity in the genome of an organism.

In yet another embodiment, the invention is a positionally addressable array comprising a plurality of different substances, selected from the group consisting of proteins, molecules comprising functional domains of said proteins, whole cells, and protein-containing cellular material, on a solid support, with each different substance being at a different position on the solid support, wherein the solid support is selected from the group consisting of ceramics, amorphous silicon carbide, castable oxides, polyimides, polymethylmethacrylates, polystyrenes and silicone elastomers.

In still another embodiment, the invention is a positionally addressable array comprising a plurality of different substances, selected from the group consisting of proteins, molecules comprising functional domains of said proteins, whole cells, and protein-containing cellular material, on a solid support, with each different substance being at a different position on the solid support, wherein the plurality of different substances are attached to the solid support via a 3-glycidooxypropyltrimethoxysilane linker.

In another embodiment, the invention is an array comprising a plurality of wells on the surface of a solid support wherein the density of the wells is at least 100 wells/cm².

The present invention also relates to a method of making a positionally addressable array comprising a plurality of wells on the surface of a solid support comprising the step of casting an array from a microfabricated mold designed to produce a density of greater than 100 wells/cm² on a solid surface. In another embodiment, the invention is a method of making a positionally addressable array comprising a plurality of wells on the surface of a solid support comprising the steps of casting a secondary mold from a microfabricated mold designed to produce a density of wells on a solid surface of greater than 100 wells/cm² and casting at least one array from the secondary mold.

In yet another embodiment, the invention is a method of using a positionally addressable array comprising a plurality of different substances, selected from the group consisting of proteins, molecules comprising functional domains of said proteins, whole cells, and protein-containing cellular material, on a solid support, with each different substance being at a different position on the solid support, wherein the plurality of different

substances consists of at least 100 different substances per cm², comprising the steps of contacting a probe with the array, and detecting protein/probe interaction.

In still another embodiment, the invention is a method of using a positionally addressable array comprising a plurality of different proteins, or molecules comprising functional domains of said proteins, on a solid support, with each different protein or molecule being at a different position on the solid support, wherein the plurality of proteins and molecules consists of at least 50% of all expressed proteins with the same type of biological activity in the genome of an organism, comprising the steps of contacting a probe with the array, and detecting protein/probe interaction.

In another embodiment, the invention is a method of using a positionally addressable array comprising a plurality of different substances, selected from the group consisting of proteins, molecules comprising functional domains of said proteins, whole cells, and protein-containing cellular material, on a solid support, with each different substance being at a different position on the solid support, wherein the solid support is selected from the group consisting of ceramics, amorphous silicon carbide, castable oxides, polyimides, polymethylmethacrylates, polystyrenes and silicone elastomers, comprising the steps of contacting a probe with the array, and detecting protein/probe interaction.

In yet another embodiment, the invention is a method of using a positionally addressable array comprising a plurality of different substances, selected from the group consisting of proteins, molecules comprising functional domains of said proteins, whole cells, and protein-containing cellular material, on a solid support, with each different substance being at a different position on the solid support, wherein the plurality of different substances are attached to the solid support via a 3-glycidooxypropyltrimethoxysilane linker, comprising the steps of contacting a probe with the array, and detecting protein/probe interaction.

In still another embodiment, the invention is a method of using a positionally addressable array comprising the steps of depositing a plurality of different substances, selected from the group consisting of proteins, molecules comprising functional domains of said proteins, whole cells, and protein-containing cellular material, on a solid support, with each different substance being at a different position on the solid support, wherein the plurality of different substances consists of at least 100 different substances per cm², contacting a probe with the array, and detecting protein/probe interaction.

In a specific embodiment, the invention is a method of using a positionally addressable array comprising the steps of depositing a plurality of different substances, selected from the group consisting of proteins, molecules comprising functional domains of said proteins, whole cells, and protein-containing cellular material, on a solid support, with

each different substance being at a different position on the solid support, wherein the plurality of different substances consists of at least 100 different substances per cm^2, and wherein the solid support is a glass slide, contacting a probe with the array, and detecting protein/probe interaction.

5   In another embodiment, the invention is a method of using a positionally addressable array comprising the steps of depositing a plurality of different proteins, or molecules comprising functional domains of said proteins, on a solid support, with each different protein or molecule being at a different position on the solid support, wherein the plurality of different proteins or molecules consists of at least 50% of all expressed proteins

10   with the same type of biological activity in the genome of an organism, contacting a probe with the array, and detecting protein/probe interaction.

In another embodiment, the invention is a method of using a positionally addressable array comprising the steps of depositing a plurality of different proteins, or molecules comprising functional domains of said proteins, on a solid support, with each

15   different protein or molecule being at a different position on the solid support, wherein the plurality of different proteins or molecules consists of at least 50% of all expressed proteins with the same type of biological activity in the genome of an organism, and wherein the solid support is a glass slide, contacting a probe with the array, and detecting protein/probe interaction.

20   In another embodiment, the invention is a method of making a positionally addressable array comprising the steps of casting an array from a microfabricated mold designed to produce a density of wells on a solid surface of greater than 100 wells/cm^2 and depositing in the wells a plurality of different substances, selected from the group consisting of proteins, molecules comprising functional domains of said proteins, whole cells, and

25   protein-containing cellular material, on a solid support, with each different substances being in a different well on the solid support.

In another embodiment, the invention is a method of making a positionally addressable array comprising the steps of casting a secondary mold from a microfabricated mold designed to produce a density of wells on a solid surface of greater than 100

30   wells/cm^2, casting at least one array from the secondary mold, and depositing in the wells a plurality of different substances, selected from the group consisting of proteins, molecules comprising functional domains of said proteins, whole cells, and protein-containing cellular material, not attached to a solid support, with each different substances being in a different well.

35   In yet another embodiment, the invention is a method of making a positionally addressable array comprising the steps of casting a secondary mold from a microfabricated

mold designed to produce a density of wells on a solid surface of greater than 100 wells/cm², casting at least one array from the secondary mold, and depositing in the wells a plurality of different substances, selected from the group consisting of proteins, molecules comprising functional domains of said proteins, whole cells, and protein-containing cellular

5   material, with each different substance being in a different well.

A.   Definitions

10   As used in this application, "protein" refers to a full-length protein, portion of a protein, or peptide. Proteins can be prepared from recombinant overexpression in an organism, preferably bacteria, yeast, insect cells or mammalian cells, or produced via fragmentation of larger proteins, or chemically synthesized.

As used in this application, "functional domain" is a domain of a protein which is

15   necessary and sufficient to give a desired functional activity. Examples of functional domains include, *inter alia*, domains which exhibit kinase, protease, phosphatase, glycosidase, acetylase, transferase, or other enzymatic activity. Other examples of functional domains include those domains which exhibit binding activity towards DNA, RNA, protein, hormone, ligand or antigen.

20   As used in this application, "probe" refers to any chemical reagent which binds to a nucleic acid (*e.g.*, DNA or RNA) or protein. Examples of probes include, *inter alia*, other proteins, peptides, oligonucleotides, polynucleotides, DNA, RNA, small molecule substrates and inhibitors, drug candidates, receptors, antigens, hormones, steroids, phospholipids, antibodies, cofactors, cytokines, glutathione, immunoglobulin domains,

25   carbohydrates, maltose, nickel, dihydrotrypsin, and biotin.

Each protein or probe on a chip is preferably located at a known, predetermined position on the solid support such that the identity of each protein or probe can be determined from its position on the solid support. Further, the proteins and probes form a positionally addressable array on a solid support.

30

IV.   Brief Description of the Drawings

Figure 1a. Using the depicted recombination strategy, 119 yeast protein kinases were cloned in a high copy *URA3* expression vector (pEGKG) that produces GST fusion

35   proteins under the control of the galactose-inducible *GAL10* promoter. GST::kinase

constructs were rescued into *E. coli*, and sequences at the 5'-end of each construct were determined. The whole procedure was repeated when mutations were discovered.

Figure 1b. Immunoblots of GST::kinase fusion proteins purified as described.
5 From three attempts, 106 kinase proteins were purified. In spite of repeated attempts, the last 14 of 119 GST fusions were undetectable by immunoblotting analysis, (*e.g.*, Mps1 in the lane labeled with star).

Figure 2a. The protein chips used in the kinase study were produced according to
10 the following process, schematically depicted. The polydimethylsiloxane (PDMS) was poured over an acrylic master mold. After curing, the chip containing the wells was peeled away and mounted on a glass slide. Next, the surface of the chip was derivatized and proteins were then attached to the wells. Wells were first blocked with 1% BSA, after which kinase, ^{33}P-γ-ATP, and buffer were added. After incubation for 30 minutes at 30°C,
15 the protein chips were washed extensively, and exposed to both X-ray film and a Molecular Dynamics PhosphorImager, which has a resolution of 50 μm and is quantitative. For twelve substrates, each kinase assay was repeated at least twice; for the remaining five substrates, the assays were performed once.

20 **Figure 2b.** An enlarged picture of a protein chip.

Figure 3. Protein chip and kinase assay results. Position I9 on every chip indicates the signal of negative control. Mps1 at position B4 showed strong kinase activities in all 12 kinase reactions, although no visible signal could be detected on a western blot (Figure 1b).
25

. **Figure 4a.** Quantitative analysis of protein kinase reactions. Kinase activities were determined using a Molecular Dynamics PhosphorImager, and the data were exported into an Excel spreadsheet. The kinase signals were then transformed into fold increases by normalizing the data against negative control. Signals of 119 kinases in four reactions are
30 shown in log scale. The fold increases ranges from 1 to 1000 fold.

Figure 4b. To determine substrate specificity, specificity index (SI) was calculated using the following formula: $SI_{ir} = F_{ir} / [(F_{i1} + F_{i2} + \cdots \cdots + F_{ir}) / r]$, where *i* represents the identity of the kinase used, *r* represents the identity of the substrate, and F_{ir} represents the
35 fold increase of a kinase *i* on substrate *r* compared with GST alone. Several examples of kinase specificity are shown when SI is greater than three.

Figure 5a. Phylogenetic tree derived from the kinase core domain multiple sequence alignment, illustrating the correlation between functional specificity and amino sequences of the poly(Tyr-Glu) kinases. Kinases that can use poly(Thr-Glu) as a substrate often map to specific regions on a sequence comparison dendrogram. The kinases that

5 efficiently phosphorylate poly(Tyr-Glu) are indicated by shading; two kinases that weakly use this substrate are indicated in boxes. Rad53 and Ste7, which could not phosphorylate poly(Tyr-Glu), are indicated by asterisks. As shown, 70% of these kinases lie in four sequence groups (circled).

10 **Figure 5b.** Structure of the rabbit muscle phosphorylase kinase (PHK)28. The positions of three basic residues and a methionine (Met) residue, which are preferentially found in kinases that can use poly(Tyr-Glu) as a substrate, are indicated. The asparagine (Asp) residue is usually found in kinases that do not use poly(Tyr-Glu).

15 **Figure 6.** Cross sectional views of lithographic steps in a process of making protein chips.

     a. A silicon wafer with two layers of silicon on either side of an oxide layer.

     b. The silicon wafer with a resistant mask layer on top.

     c. The etching process removes silicon where the surface is unprotected by the

20 resistant mask. The depth of the etching is controlled by the position of the oxide layer, *i.e.*, the etching process does not remove the oxide layer.

     d. The mask layer is removed, leaving the etched silicon wafer.

     e. The protein chip material is applied to the mold.

     f. After curing, the protein chip is removed from the mold. The protein chip has an

25 image that is the negative of the mold.

Figure 7. Kinase/inhibitor assays on a protein chip. A human protein kinase A (PKA), a human map kinase (MAPK), three yeast PKA homologs (TPK1, TPK2 and TPK3), and two other yeast protein kinases (HSL1 and RCK1) were tested against two

30 substrates (*i.e.*, a protein substrate for PKA and a commonly used kinase substrate, MBP) using different concentrations of a specific human PKA inhibitor, PKIα, or a MAPK inhibitor, SB202190. As shown in the figure, PKIα can specifically inhibit PKA activities using both peptide and MBP as substrates. However, SB202190 did not show any inhibitory effect on PKA activity. It is also interesting to note that PKIα did not inhibit the

35 three yeast PKA homologs (TPK1, TPK2, TPK3) or the other two yeast protein kinases tested, HSL1 and RCK1.

- 9 -

V. Detailed Description of the Invention

The invention is directed to protein chips, *i.e.*, positionally addressable arrays of proteins on a solid support, useful for the large-scale study of protein function, wherein the protein chip contains densely packed reaction wells. A positionally addressable array provides a configuration such that each probe or protein of interest is located at a known, predetermined position on the solid support such that the identity of each probe or protein can be determined from its position on the array. The invention is also directed to methods of using protein chips to assay the presence, amount, and/or functionality of proteins present in at least one sample. The invention also is directed to methods of using the protein chips for high density and small volume chemical reactions. Also, the invention is directed to polymers useful as protein chip substrates and methods of making protein chips. The invention further relates to compounds useful for the derivatization of protein chip substrate.

In one embodiment, the invention is a positionally addressable array comprising a plurality of different substances, selected from the group consisting of proteins, molecules comprising functional domains of said proteins, whole cells, and protein-containing cellular material, on a solid support, with each different substance being at a different position on the solid support, wherein the plurality of different substances consists of at least 100 different substances per cm^2. In one embodiment, said plurality of different substances consists of between 100 and 1000 different substances per cm^2. In another embodiment, said plurality of different substances consists of between 1000 and 10,000 different substances per cm^2. In another embodiment, said plurality of different substances consists of between 10,000 and 100,000 different substances per cm^2. In yet another embodiment, said plurality of different substances consists of between 100,000 and 1,000,000 different substances per cm^2. In yet another embodiment, said plurality of different substances consists of between 1,000,000 and 10,000,000 different substances per cm^2. In yet another embodiment, said plurality of different substances consists of between 10,000,000 and 25,000,000 different substances per cm^2. In yet another embodiment, said plurality of different substances consists of at least 25,000,000 different substances per cm^2. In yet another embodiment, said plurality of different substances consists of at least 10,000,000,000 different substances per cm^2. In yet another embodiment, said plurality of different substances consists of at least 10,000,000,000,000 different substances per cm^2.

In another embodiment, the invention is a positionally addressable array comprising a plurality of different substances, selected from the group consisting of proteins, molecules comprising functional domains of said proteins, whole cells, and protein-containing cellular

material, on a solid support, with each different substance being at a different position on the solid support, wherein the plurality of different substances consists of at least 100 different substances per cm^2, and wherein the solid support is a glass slide.

In another embodiment, the invention is a positionally addressable array comprising a plurality of different substances, selected from the group consisting of proteins, molecules comprising functional domains of said proteins, whole cells, and protein-containing cellular material, on a solid support, with each different substance being at a different position on the solid support, wherein the plurality of different substances consists of about 30 to 100 different substances per cm^2. In a specific embodiment, said plurality of different substances consists of 30 different substances per cm^2. In a particular embodiment, said plurality of different substances consists of between 30 and 50 different substances per cm^2. In another particular embodiment, said plurality of different substances consists of between 50 and 100 different substances per cm^2.

In various specific embodiments, the invention is a positionally addressable array comprising a plurality of different proteins, or molecules comprising functional domains of said proteins, on a solid support, with each different protein or molecule being at a different position on the solid support, wherein the plurality of different proteins or molecules consists of at least 50%, 75%, 90%, or 95% of all expressed proteins with the same type of biological activity in the genome of an organism. For example, such organism can be eukaryotic or prokaryotic, and is preferably a mammal, a human or non-human animal, primate, mouse, rat, cat, dog, horse, cow, chicken, fungus such as yeast, *Drosophila*, *C. elegans*, etc. Such type of biological activity of interest can be, but is not limited to, enzymatic activity (*e.g.*, kinase activity, protease activity, phosphatase activity, glycosidase, acetylase activity, and other chemical group transferring enzymatic activity), nucleic acid binding, hormone binding, etc.

A.   **Production of Protein Chips**

The protein chips with densities of wells in an array of the present invention are preferably cast from master molds which have been stamped, milled, or etched using conventional microfabrication or microlithographic techniques. Preferably conventional microlithographic techniques and materials are utilized in the production of the master molds. Once a master mold has been produced, the master mold may then be used directly to mold the protein chips per se. Alternatively, secondary or tertiary molds can be cast from the master mold and the protein chips cast from these secondary or tertiary molds.

The master mold can be made from any material that is suitable for microfabrication or microlithography, with silicon, glass, quartz, polyimides, and polymethylmethacrylate (Lucite) being preferred. For microlithography, the preferred material is silicon wafers.

Once the appropriate master, secondary, or tertiary mold has been produced, the
5 protein chip is cast. The protein chip can be cast in any solid support that is suitable for casting, including either porous or non-porous solid supports. Ceramics, amorphous silicon carbide, castable oxides that produce casts of SiO_2 when cured, polyimides, polymethylmethacrylates, and polystyrenes are preferred solid supports, with silicone elastomeric materials being most preferred. Of the silicone elastomeric materials,
10 polydimethylsiloxane (PDMS) is the most preferred solid support. An advantage of silicone elastomeric materials is the ease with which they are removed from the mold due to their flexible nature.

Figure 6 illustrates an example of one method useful for realizing high-density arrays of wells on protein chips according to this invention. A silicon wafer with an oxide
15 layer sandwiched between layers of silicon is provided (Figure 6a). Known as silicon-on-insulator or SOI wafers, these wafers are commonly available from wafer supply companies (e.g., Belle Mead Research, Belle Mead, NJ, and Virginia Semiconductor, Fredericksburg, VA).

The silicon wafer is then patterned and etched via an etch process (Figures 6b-d).
20 The buried oxide layer acts as a very effective etch stop and results in highly uniform etch depth across the wafer. Etch depth is independent of the etch process and merely is determined by the thickness of the top silicon layer.

A wet chemical etch process (e.g., using KOH or tetra-methyl hydrazine (TMAH)) can be utilized. However, this technique is slightly more dependent on the crystal
25 orientation of the silicon wafer. Thus, a technique using a rarefied gas (typically SF_6) in a reactive ion etch (RIE) is preferred. RIE etching techniques are capable of realizing highly anisotropic wells in silicon that are independent of the crystal orientation of the silicon wafer. The references G. Kovacs, Micromachined Transducers Sourcebook, Academic Press (1998) and M. Madou, Fundamentals of Microfabrication, CRC Press (1997) provide
30 background on etching techniques.

Both types of microlithography can be utilized on a single chip to obtain the desired combination of well shapes. Wet-chemical etching is an isotropic process which gives U-shaped wells, while RIE is an anisotropic process which gives square bottomed wells.

After etching the wafer to realize a master mold, it can be used to cast protein chips
35 (Figures 6e-f). These structures can be the protein chips or themselves be secondary or tertiary molds from which additional casting of protein chips occurs.

- 12 -

Thus, in one embodiment, a method of making a positionally addressable array, comprising a plurality of wells on the surface of a solid support, comprises casting an array from a microfabricated mold designed to produce a density of wells on a solid surface of greater than 100 wells/cm^2. In another embodiment, a method of making a positionally

5   addressable array, comprising a plurality of wells on the surface of a solid support, comprises casting a secondary mold from said microfabricated mold designed to produce a density of wells on a solid surface of greater than 100 wells/cm^2 and casting at least one array from the secondary mold. In yet another embodiment, a method of making a positionally addressable array comprises covering the mold with a liquid cast material, and

10   curing the cast material until the cast is solid. The liquid cast material is preferably silicone elastomer, most preferably polydimethylsiloxane. Into any of these positionally addressable arrays, a plurality of different substances, selected from the group consisting of proteins, molecules comprising functional domains of said proteins, whole cells, and protein-containing cellular material, can be deposited such that each different substance is found in

15   a different well on the solid support.

B.   **Features of Protein Chips**

The protein chips of the present invention are not limited in their physical

20   dimensions and may have any dimensions that are convenient. For the sake of compatibility with current laboratory apparatus, protein chips the size of a standard microscope slide or smaller are preferred. Most preferred are protein chips sized such that two chips fit on a microscope slide. Also preferred are protein chips sized to fit into the sample chamber of a mass spectrometer.

25   The wells in the protein chips of the present invention may have any shape such as rectangular, square, or oval, with circular being preferred. The wells in the protein chips may have square or round bottoms, V-shaped bottoms, or U-shaped bottoms. Square bottoms are slightly preferred because the preferred reactive ion etch (RIE) process, which is anisotropic, provides square-bottomed wells. The shape of the well bottoms need not be

30   uniform on a particular chip, but may vary as required by the particular assay being carried out on the chip.

The wells in the protein chips of the present invention may have any width-to-depth ratio, with ratios of width-to-depth between about 10:1 and about 1:10 being preferred. The wells in the protein chips of the present invention may have any volume, with wells having

35   volumes of between 1 pl and 5 µl preferred and wells having volumes of between 1 nl and 1 µl being more preferred. The most preferred volume for a well is between 100 nl and 300

nl. For protein chips with very high densities of wells, the preferred volume of a well is between 10 pl and 100 nl.

The protein chips of the invention can have a wide variety of density of wells/cm^2. The preferred density of wells is between about 25 wells/cm^2 and about 10,000,000,000,000

5   wells/cm^2. Densities of wells on protein chips cast from master molds of laser milled Lucite are generally between 1 well/cm^2 and 2,500 wells/cm^2. Appropriate milling tools produce wells as small as 100 μm in diameter and 100 μm apart. Protein chips cast from master mold etched by wet-chemical microlithographic techniques have densities of wells generally between 50 wells/cm^2 and 10,000,000,000 wells/cm^2. Wet-chemical etching can

10   produce wells that are 10 μm deep and 10 μm apart, which in turn produces wells that are less than 10 μm in diameter. Protein chips cast from master mold etched by RIE microlithographic techniques have densities of wells generally between 100 wells/cm^2 and 25,000,000 wells/cm^2. RIE in combination with optical lithography can produce wells that are 500 nm in diameter and 500 nm apart. Use of electron beam lithography in combination

15   with RIE can produce wells 50 nm in diameter and 50 nm apart. Wells of this size and with equivalent spacing produces protein chips with densities of wells 10,000,000,000,000 wells/cm^2. Preferably, RIE is used to produce wells of 20 μm in diameter and 20 μm apart. Wells of this size that are equivalently spaced will result in densities of 25,000,000 wells/cm^2.

20   The microfabrication and microlithographic techniques described above have been used successfully to wet-chemically etch silicon wafers with well sizes of 560 μm or 280 μm with spacing of about 1 mm. This combination of wells and spacing produces arrays of about 410,000 wells/cm^2 and about 610,000 wells/cm^2, respectively. When well size and spacing are equivalent, protein chips with about 3.19 million wells/cm^2 and 12.75 million

25   wells/cm^2 are produced.

In one embodiment, the array comprises a plurality of wells on the surface of a solid support wherein the density of wells is at least 100 wells/cm^2. In another embodiment, said density of wells is between 100 and 1000 wells/cm^2. In another embodiment, said density of wells is between 1000 and 10,000 wells/cm^2. In another embodiment, said density of

30   wells is between 10,000 and 100,000 wells/cm^2. In yet another embodiment, said density of wells is between 100,000 and 1,000,000 wells/cm^2. In yet another embodiment, said density of wells is between 1,000,000 and 10,000,000 wells/cm^2. In yet another embodiment, said density of wells is between 10,000,000 and 25,000,000 wells/cm^2. In yet another embodiment, said density of wells is at least 25,000,000 wells/cm^2. In yet another

35   embodiment, said density of wells is at least 10,000,000,000 wells/cm^2. In yet another embodiment, said density of wells is at least 10,000,000,000,000 wells/cm^2.

NY2 - 1102584.9

C.   Utilization of Protein Chips

        In one embodiment, the present invention provides a protein chip comprising a flat surface, such as, but not limited to, glass slides. Dense protein arrays can be produced on,
5 for example, glass slides, such that chemical reactions and assays can be conducted, thus allowing large-scale parallel analysis of the presence, amount, and/or functionality of proteins (*e.g.*, protein kinases). Proteins or probes are bound covalently or non-covalently to the flat surface of the solid support. The proteins or probes can be bound directly to the flat surface of the solid support, or can be attached to the solid support through a linker
10 molecule or compound. The linker can be any molecule or compound that derivatizes the surface of the solid support to facilitate the attachment of proteins or probes to the surface of the solid support. The linker may covalently or non-covalently bind the proteins or probes to the surface of the solid support. In addition, the linker can be an inorganic or organic molecule. Preferred linkers are compounds with free amines. Most preferred
15 among linkers is 3-glycidooxypropyltrimethoxysilane (GPTS).
        In another embodiment, the protein chips of the present invention have several advantages over flat surface arrays. Namely, the use of wells eliminates or reduces the likelihood of cross-contamination with respect to the contents of the wells. Another advantage over flat surfaces is increased signal-to-noise ratios. Wells allow the use of
20 larger volumes of reaction solution in a denser configuration, and therefore greater signal is possible. Furthermore, wells decrease the rate of evaporation of the reaction solution from the chip as compared to flat surface arrays, thus allowing longer reaction times.
        Another advantage of wells over flat surfaces is that the use of wells permit association studies using a fixed, limited amount of probe for each well on the chip,
25 whereas the use of flat surfaces usually involves indiscriminate probe application across the whole substrate. When a probe in a mixture of probes has a high affinity, but low specificity, the indiscriminate application of the probe mixture across the substrate will saturate many of the proteins with the high affinity probe. This saturation effectively limits the detection of other probes in the mixture. By using wells, a limited amount of a probe
30 can be applied to individual wells on the chip. Thus, the amount of the probe applied to individual proteins can be controlled, and the probe can be different for different proteins (situated in different wells).
        Once a protein chip is produced as described above, it can be used to conduct assays and other chemical reactions. For assays, proteins or probes will generally be placed in the
35 wells. The presence or absence of proteins or probes will be detected by the application of probes or proteins, respectively, to the protein chip. The protein-probe interaction can be

NY2 - 1102584.9

visualized using a variety of techniques known in the art, some of which are discussed below.

Proteins useful in this invention can be fusion proteins, in which a defined domain is attached to one of a variety of natural proteins, or can be intact non-fusion proteins.

5   In another embodiment, protein-containing cellular material, such as but not limited to vesicles, endosomes, subcellular organelles, and membrane fragments, can be placed on the protein chip (*e.g.*, in wells). In another embodiment, a whole cell is placed on the protein chip (*e.g.*, in wells). In a further embodiment, the protein, protein-containing cellular material, or whole cell is attached to the solid support of the protein chip.

10   The protein can be purified prior to placement on the protein chip or can be purified during placement on the chip via the use of reagents that bind to particular proteins, which have been previously placed on the protein chip. Partially purified protein-containing cellular material or cells can be obtained by standard techniques (*e.g.*, affinity or column chromatography) or by isolating centrifugation samples (*e.g.*, P1 or P2 fractions).

15   Furthermore, proteins, protein-containing cellular material, or cells can be embedded in artificial or natural membranes prior to or at the time of placement on the protein chip. In another embodiment, proteins, protein-containing cellular material, or cells can be embedded in extracellular matrix component(s) (*e.g.*, collagen or basal lamina) prior to or at the time of placement on the protein chip. The proteins of the invention can be in solution,

20 or bound to the surface of the solid support (*e.g.*, in a well, or on a flat surface), or bound to a substrate (*e.g.*, bead) placed in a well of the solid support.

The placement of proteins or probes in the wells can be accomplished by using any dispensing means, such as bubble jet or ink jet printer heads. A micropipette dispenser is preferred. The placement of proteins or probes can either be conducted manually or the

25 process can be automated through the use of a computer connected to a machine.

Since the wells are self-contained, the proteins or probes need not be attached or bound to the surface of the solid support, but rather the proteins or probes can simply be placed in the wells, or bound to a substrate (*e.g.*, bead) that is placed in the wells. Other substrates include, but are not limited to, nitrocellulose particles, glass beads, plastic beads,

30 magnetic particles, and latex particles. Alternatively, the proteins or probes are bound covalently or non-covalently to the surface of the solid support in the wells. The proteins or probes can be bound directly to the surface of the solid support (in the well), or can be attached to the solid support through a linker molecule or compound. The linker can be any molecule or compound that derivatizes the surface of the solid support to facilitate the

35 attachment of proteins or probes to the surface of the solid support. The linker may covalently bind the proteins or probes to the surface of the solid support or the linker may

bind via non-covalent interactions. In addition, the linker can be an inorganic or organic molecule. Preferred linkers are compounds with free amines. Most preferred among linkers is 3-glycidooxypropyltrimethoxysilane (GPTS).

Proteins or probes which are non-covalently bound to the well surface may utilize a variety of molecular interactions to accomplish attachment to the well surface such as, for example, hydrogen bonding, van der Waals bonding, electrostatic, or metal-chelate coordinate bonding. Further, DNA-DNA, DNA-RNA and receptor-ligand interactions are types of interactions that utilize non-covalent binding. Examples of receptor-ligand interactions include interactions between antibodies and antigens, DNA-binding proteins and DNA, enzyme and substrate, avidin (or streptavidin) and biotin (or biotinylated molecules), and interactions between lipid-binding proteins and phospholipid membranes or vesicles. For example, proteins can be expressed with fusion protein domains that have affinities for a substrate that is attached to the surface of the well. Suitable substrates for fusion protein binding include trypsin/anhydrotrypsin, glutathione, immunoglobulin domains, maltose, nickel, or biotin and its derivatives, which bind to bovine pancreatic trypsin inhibitor, glutathione-S-transferase, antigen, maltose binding protein, poly-histidine (*e.g.*, HisX6 tag), and avidin/streptavidin, respectively.

D.   Assays on Protein Chips

In one embodiment, the protein chips are used in assays by using standard enzymatic assays that produce chemiluminescence or fluorescence. Detection of various proteins and molecular modifications can be accomplished using, for example, photoluminescence, fluorescence using non-protein substrates, enzymatic color development, mass spectroscopic signature markers, and amplification (*e.g.*, by PCR) of oligonucleotide tags. Thus, protein/probe interaction can be detected by, *inter alia*, chemiluminescence, fluorescence, radiolabeling, or atomic force microscopy. Probes binding to specific elements in the array can also be identified by direct mass spectrometry. For example, probes released into solution by non-degradative methods, which dissociate the probes from the array elements, can be identified by mass spectrometry (*see, e.g.*, WO 98/59361). In another example, peptides or other compounds released into solution by enzymatic digests of the array elements can be identified by mass spectrometry.

The types of assays fall into several general categories. As a first example, each well on the array is exposed to a single probe whose binding is detected and quantified. The results of these assays are visualized by methods including, but not limited to: 1) using radioactively labeled ligand followed by autoradiography and/or phosphoimager analysis;

2) binding of hapten, which is then detected by a fluorescently labeled or enzymatically labeled antibody or high affinity hapten ligand such as biotin or streptavidin; 3) mass spectrometry; 4) atomic force microscopy; 5) fluorescent polarization methods; 6) rolling circle amplification-detection methods (Hatch et al., 1999, "Rolling circle amplification of

5   DNA immobilized on solid surfaces and its application to multiplex mutation detection", Genet. Anal. 15(2):35-40); 7) competitive PCR (Fini et al., 1999, "Development of a chemiluminescence competitive PCR for the detection and quantification of parvovirus B19 DNA using a microplate luminometer", Clin Chem. 45(9):1391-6; Kruse et al., 1999, "Detection and quantitative measurement of transforming growth factor-beta1 (TGF-beta1)

10 gene expression using a semi-nested competitive PCR assay", Cytokine 11(2):179-85; Guenthner and Hart, 1998, "Quantitative, competitive PCR assay for HIV-1 using a microplate-based detection system", Biotechniques 24(5):810-6); 8) colorimetric procedures; and 9) biological assays, e.g., for virus titers.

As a second example, each well on the array is exposed to multiple probes

15 concurrently, including pooling of probes from several sources, whose binding is detected and quantified. The results of these assays are visualized by methods including, but not limited to: 1) mass spectrometry; 2) atomic force microscopy; 3) infrared red or fluorescently labeled compounds or proteins; 4) amplifiable oligonucleotides, peptides or molecular mass labels; and 5) by stimulation or inhibition of the protein's enzymatic

20 activity. Information is gleaned from mixtures of probes because of the positionally addressable nature of the arrays of the present invention, i.e., through the placement of defined proteins at known positions on the protein chip, information about to what the bound probe binds is known. If so desired, positions on the array that demonstrate binding can then be probed with individual probes to identify the specific interaction of interest.

25 Useful information also can be obtained, for example, by incubating a protein chip with cell extracts, wherein each well on the chip contains a reaction mix to assay an enzymatic activity of interest, and wherein a plurality of different enzymatic and/or substrate activities are assayed, and thereby identifying and measuring the cellular repertoire of particular enzymatic activities. Similarly, the protein chip can be incubated

30 with whole cells or preparations of plasma membranes to assay, for example, for expression of membrane-associated proteins or molecules, or binding properties of cell surface proteins or molecules. Cells, markers on a cell, or substances secreted by a cell that bind to particular locations on the protein chip can be detected using techniques known in the art. For example, protein chips containing arrays of antigens can be screened with B-cells or T-

35 cells, wherein the antigens are selected from the group consisting of synthetic antigens, tissue-specific antigens, disease-specific antigens, antigens of pathogens, and antigens of

- 18 -

autologous tissues. The antigen or antigenic determinant recognized by the lymphocytes can be determined by establishing at what position on the array activation of the cells by antigen occurs. Lymphocyte activation can be assayed by various means including, but not limited to, detecting antibody synthesis, detecting or measuring incorporation of ^{3}H-

5   thymidine, probing of cell surface molecules with labeled antibodies to identify molecules induced or suppressed by antigen recognition and activation (*e.g.*, IgD, C3b receptor, IL-2 receptor, transferrin receptor, membrane class II MHC molecules, CD23, CD38, PCA-1 molecules, HLA-DR), and identify expressed and/or secreted cytokines.

In another example, mitogens for a specific cell-type can be determined by

10   incubating the cells with protein chips containing arrays of putative mitogens, comprising the steps of contacting a positionally addressable array with a population of cells; said array comprising a plurality of different substances, selected from the group consisting of proteins, molecules comprising functional domains of said proteins, whole cells, and protein-containing cellular material, on a solid support, with each different substance being

15   at a different position on the solid support, wherein the density of different substances is at least 100 different substances per cm^2; and detecting positions on the solid support where mitogenic activity is induced in a cell. Cell division can be assayed by, for example, detecting or measuring incorporation of ^{3}H-thymidine by a cell. Cells can be of the same cell type (*i.e.*, a homogeneous population) or can be of different cell types.

20   In yet another example, cellular uptake and/or processing of proteins on the protein chips can be assayed by, for example, using radioactively labeled protein substrates and measuring either a decrease in radioactive substrate concentration or uptake of radioactive substrate by the cells.   These assays can be used for either diagnostic or therapeutic purposes. One of ordinary skill in the art can appreciate many appropriate assays for

25   detecting various types of cellular interactions.

Thus, use of several classes of probes (*e.g.*, known mixtures of probes, cellular extracts, subcellular organelles, cell membrane preparations, whole cells, etc.) can provide for large-scale or exhaustive analysis of cellular activities. In particular, one or several screens can form the basis of identifying a "footprint" of the cell type or physiological state

30   of a cell, tissue, organ or system. For example, different cell types (either morphological or functional) can be differentiated by the pattern of cellular activities or expression determined by the protein chip. This approach also can be used to determine, for example, different stages of the cell cycle, disease states, altered physiologic states (*e.g.*, hypoxia), physiological state before or after treatment (*e.g.*, drug treatment), metabolic state, stage of

35   differentiation or development, response to environmental stimuli (*e.g.*, light, heat), cell-cell

interactions, cell-specific gene and/or protein expression, and disease-specific gene and/or protein expression.

Enzymatic reactions can be performed and enzymatic activity measured using the protein chips of the present invention. In a specific embodiment, compounds that modulate the enzymatic activity of a protein or proteins on a chip can be identified. For example, changes in the level of enzymatic activity are detected and quantified by incubation of a compound or mixture of compounds with an enzymatic reaction mixture in wells of the protein chip, wherein a signal is produced (e.g., from substrate that becomes fluorescent upon enzymatic activity). Differences between the presence and absence of the compound are noted. Furthermore, the differences in effects of compounds on enzymatic activities of different proteins are readily detected by comparing their relative effect on samples within the protein chips and between chips.

The variety of strategies of using the high density protein chips of the present invention, detailed above, can be used to determine various physical and functional characteristics of proteins. For example, the protein chips can be used to assess the presence and amount of protein present by probing with an antibody. In one embodiment, a polydimethylsiloxane (PDMS) chip of GST fusion proteins can be probed to determine the presence of a protein and/or its level of activity. The protein can be detected using standard detection assays such as luminescence, chemiluminescence, fluorescence or chemifluorescence. For example, a primary antibody to the protein of interest is recognized by a fluorescently labeled secondary antibody, which is then measured with an instrument (e.g., a Molecular Dynamics scanner) that excites the fluorescent product with a light source and detects the subsequent fluorescence. For greater sensitivity, a primary antibody to the protein of interest is recognized by a secondary antibody that is conjugated to an enzyme such as alkaline phosphatase or horseradish peroxidase. In the presence of a luminescent substrate (for chemiluminescence) or a fluorogenic substrate (for chemifluorescence), enzymatic cleavage yields a highly luminescent or fluorescent product which can be detected and quantified by using, for example, a Molecular Dynamics scanner. Alternatively, the signal of a fluorescently labeled secondary antibody can be amplified using an alkaline phosphatase-conjugated or horseradish peroxidase-conjugated tertiary antibody.

Identifying substrates of protein kinases, phosphatases, proteases, glycosidases, acetylases, or other group transferring enzymes can also be conducted on the protein chips of the present invention. For example, a wide variety of different probes are attached to the protein chip and assayed for their ability to act as a substrate for particular enzyme(s), e.g., assayed for their ability to be phosphorylated by protein kinases. Detection methods for

kinase activity, include, but are not limited to, the use of radioactive labels, such as ^{33}P-ATP and ^{35}S-γ-ATP, or fluorescent antibody probes that bind to phosphoamino acids. For example, whereas incorporation into a protein of radioactively labeled phosphorus indicates kinase activity in one assay, another assay can measure the release of radioactively labeled

5 phosphorus into the media, which indicates phosphatase activity. In another example, protease activity can be detected by identifying, using standard assays (e.g., mass spectrometry, fluorescently labeled antibodies to peptide fragments, or loss of fluorescence signal from a fluorescently tagged substrate), peptide fragments that are produced by protease activity and released into the media. Thus, activity of group-transferring enzymes

10 can be assayed readily using several approaches and many independent means of detection, which would be appreciated by one of ordinary skill in the art.

Protein chips can be used to identify proteins on the chip that have specific activities such as specific kinases, proteases, nucleic acid binding properties, nucleotide hydrolysis, hormone binding and DNA binding. Thus, the chip can be probed with a probe that will

15 indicate the presence of the desired activity. For example, if DNA binding is the activity of interest, the chip containing candidate DNA-binding proteins is probed with DNA.

The search for probes (natural or synthetic) that are protein or nucleic acid ligands for an array of proteins can be carried out in parallel on a protein chip. A probe can be a cell, protein-containing cellular material, protein, oligonucleotide, polynucleotide, DNA,

20 RNA, small molecule substrate, drug candidate, receptor, antigen, steroid, phospholipid, antibody, immunoglobulin domain, glutathione, maltose, nickel, dihydrotrypsin, or biotin. Alternatively, the probe can be an enzyme substrate or inhibitor. For example, the probe can be a substrate or inhibitor of an enzyme chosen from the group consisting of kinases, phosphatases, proteases, glycosidases, acetylases, and other group transferring enzymes.

25 After incubation of proteins on a chip with combinations of nucleic acid or protein probes, the bound nucleic acid or protein probes can be identified by mass spectrometry (Lakey et al., 1998, "Measuring protein-protein interactions", Curr Opin Struct Biol. 8:119-23).

The identity of target proteins from pathogens (e.g., an infectious disease agent such as a virus, bacterium, fungus, or parasite) or target proteins from abnormal cells (e.g.,

30 neoplastic cells, diseased cells, or damaged cells) that serve as antigens in the immune response of recovering or non-recovering patients can be determined by using a protein chip of the invention. For example, lymphocytes isolated from a patient can be used to screen protein chips comprising arrays of a pathogen's proteins on a protein chip. In general, these screens comprise contacting a positionally addressable array with a plurality of

35 lymphocytes, said array comprising a plurality of potential antigens on a solid support, with each different antigen being at a different position on the solid support, wherein the density

NY2 - 1102584.9

of different antigens is at least 100 different antigens per cm^2, and detecting positions on the solid support where lymphocyte activation occurs. In a specific embodiment, lymphocytes are contacted with a pathogen's proteins on an array, after which activation of B-cells or T-cells by an antigen or a mixture of antigens is assayed, thereby identifying target antigens

5   derived from a pathogen.

Alternatively, the protein chips are used to characterize an immune response by, for example, screening arrays of potential antigens to identify the targets of a patient's B-cells and/or T-cells. For example, B-cells can be incubated with an array of potential antigens (*i.e.*, molecules having antigenic determinants) to identify antigenic targets for humoral-

10   based immunity. The source of antigens can be, for example, from autologous tissues, collections of known or unknown antigens (*e.g.*, of pathogenic microorganisms), tissue-specific or disease-specific antigen collections, or synthetic antigens.

In another embodiment, lymphocytes isolated from a patient can be used to screen protein chips comprising arrays of proteins derived from a patient's own tissues. Such

15   screens can identify substrates of autoimmunity or allergy-causing proteins, and thereby diagnose autoimmunity or allergic reactions, and/or identify potential target drug candidates.

In another embodiment, the protein chips of the invention are used to identify substances that are able to activate B-cells or T-cells. For example, lymphocytes are

20   contacted with arrays of test molecules or proteins on a chip, and lymphocyte activation is assayed, thereby identifying substances that have a general ability to activate B-cells or T-cells or subpopulations of lymphocytes (*e.g.*, cytotoxic T-cells).

Induction of B-cell activation by antigen recognition can be assayed by various means including, but not limited to, detecting or measuring antibody synthesis,

25   incorporation of ^{3}H-thymidine, binding of labeled antibodies to newly expressed or suppressed cell surface molecules, and secretion of factors indicative of B-cell activation (*e.g.*, cytokines). Similarly, T-cell activation in a screen using a protein chip of the invention can be determined by various assays. For example, a chromium (^{51}Cr) release assay can detect recognition of antigen and subsequent activation of cytotoxic T-cells (*see,*

30   *e.g.*, Palladino et al., 1987, Cancer Res. 47:5074-9; Blachere et al., 1993, J. Immunotherapy 14:352-6).

The specificity of an antibody preparation can be determined through the use of a protein chip of the invention, comprising contacting a positionally addressable array with an antibody preparation, said array comprising a plurality of potential antigens on a solid

35   support, with each different antigen being at a different position on the solid support, wherein the density of different antigens is at least 100 different antigens per cm^2, and

detecting positions on the solid support where binding by an antibody in the antibody preparation occurs. The antibody preparation can be, but is not limited to, Fab fragments, antiserum, and polyclonal, monoclonal, chimeric, single chain, humanized, or synthetic antibodies. For example, an antiserum can be characterized by screening disease-specific,

5   tissue-specific, or other identified collections of antigens, and determining which antigens are recognized. In a specific embodiment, protein chip arrays having similar or related antigens are screened with monoclonal antibodies to evaluate the degree of specificity by determining to which antigens on the array a monoclonal antibody binds.

The identity of targets of specific cellular activities can be assayed by treating a

10   protein chip with complex protein mixtures, such as cell extracts, and determining protein activity. For example, a protein chip containing an array of different kinases can be contacted with a cell extract from cells treated with a compound (*e.g.*, a drug), and assayed for kinase activity. In another example, a protein chip containing an array of different kinases can be contacted with a cell extract from cells at a particular stage of cell

15   differentiation (*e.g.*, pluripotent) or from cells in a particular metabolic state (*e.g.*, mitotic), and assayed for kinase activity. The results obtained from such assays, comparing for example, cells in the presence or absence of a drug, or cells at several differentiation stages, or cells in different metabolic states, can provide information regarding the physiologic changes in the cells between the different conditions.'

20   Alternatively, the identity of targets of specific cellular activities can be assayed by treating a protein chip of the invention, containing many different proteins *(e.g.*, a peptide library), with a complex protein mixture (*e.g.,* such as a cell extract), and assaying for modifications to the proteins on the chip. For example, a protein chip containing an array of different proteins can be contacted with a cell extract from cells treated with a compound

25   (*e.g.*, a drug), and assayed for kinase, protease, glycosidase, actetylase, phosphatase, or other transferase activity, for example. In another example, a protein chip containing an array of different proteins can be contacted with a cell extract from cells at a particular stage of cell differentiation (*e.g.*, pluripotent) or from cells in a particular metabolic state (*e.g.*, mitotic). The results obtained from such assays, comparing for example, cells in the

30   presence or absence of a drug, or cells at several stages of differentiation, or cells in different metabolic states, can provide information regarding the physiologic effect on the cells under these conditions.

The protein chips are useful to identify probes that bind to specific molecules of biologic interest including, but not limited to, receptors for potential ligand molecules, virus

35   receptors, and ligands for orphan receptors.

The protein chips are also useful to detecting DNA binding or RNA binding to proteins on the protein chips, and to determine the binding specificity. In addition, particular classes of RNA-binding or DNA-binding proteins (*e.g.*, zinc-finger proteins) can be studied with the protein chips by screening arrays of these proteins with nucleic acid

5   sequences, and determining binding specificity and binding strength.

The identity of proteins exhibiting differences in function, ligand binding, or enzymatic activity of similar biological entities can be analyzed with the protein chips of the present invention. For example, differences in protein isoforms derived from different alleles are assayed for their activities relative to one another.

10   The high density protein chips can be used for drug discovery, analysis of the mode of action of a drug, drug specificity, and prediction of drug toxicity. For example, the identity of proteins that bind to a drug, and their relative affinities, can be assayed by incubating the proteins on the chip with a drug or drug candidate under different assay conditions, determining drug specificity by determining where on the array the drug bound,

15   and measuring the amount of drug bound by each different protein. Bioassays in which a biological activity is assayed, rather than binding assays, can alternatively be carried out on the same chip, or on an identical second chip. Thus, these types of assays using the protein chips of the invention are useful for studying drug specificity, predicting potential side effects of drugs, and classifying drugs. Further, protein chips of the invention are suitable

20   for screening complex libraries of drug candidates. Specifically, the proteins on the chip can be incubated with the library of drug candidates, and then the bound components can be identified, *e.g.*, by mass spectrometry, which allows for the simultaneous identification of all library components that bind preferentially to specific subsets of proteins, or bind to several, or all, of the proteins on the chip. Further, the relative affinity of the drug

25   candidates for the different proteins in the array can be determined.

Moreover, the protein chips of the present invention can be probed in the presence of potential inhibitors, catalysts, modulators, or enhancers of a previously observed interaction, enzymatic activity, or biological response. In this manner, for example, blocking of the binding of a drug, or disruption of virus or physiological effectors to

30   specific categories of proteins, can be analyzed by using a protein chip of the present invention.

The protein chips of the invention can be used to determine the effects of a drug on the modification of multiple targets by complex protein mixtures, such as for example, whole cells, cell extracts, or tissue homogenates. The net effect of a drug can be analyzed

35   by screening one or more protein chips with drug-treated cells, tissues, or extracts, which then can provide a "signature" for the drug-treated state, and when compared with the

"signature" of the untreated state, can be of predictive value with respect to, for example, potency, toxicity, and side effects. Furthermore, time-dependent effects of a drug can be assayed by, for example, adding the drug to the cell, cell extract, tissue homogenate, or whole organism, and applying the drug-treated cells or extracts to a protein chip at various 5 timepoints of the treatment.

Screening of phage display libraries can be performed by incubating a library with the protein chips of the present invention. Binding of positive clones can be determined by various methods known in the art (*e.g.*, mass spectrometry), thereby identifying clones of interest, after which the DNA encoding the clones of interest can be identified by standard 10 methods (*see, e.g.*, Ames et al., 1995, J. Immunol. Methods 184:177-86; Kettleborough et al., 1994, Eur. J. Immunol. 24:952-8; Persic et al., 1997, Gene 187:9-18). In this manner, the chips are useful to select for cells having surface components that bind to specific proteins on the chip. Alternatively, a phage display library can be attached to the chip, such that a positionally addressable array of the library is created, after which the array can be 15 screened repeatedly with different mixtures of probes.

The invention also provides kits for carrying out the assay regimens of the invention. In a specific embodiment, kits of the invention comprise one or more arrays of the invention. Such kits may further comprise, in one or more containers, reagents useful for assaying biological activity of a protein or molecule, reagents useful for assaying 20 interaction of a probe and a protein or molecule, reagents useful for assaying the biological activity of a protein or molecule having a biological activity of interest, and/or one or more probes, proteins or other molecules. The reagents useful for assaying biological activity of a protein or molecule, or assaying interactions between a probe and a protein or molecule, can be contained in each well or selected wells on the protein chip. Such reagents can be in 25 solution or in solid form. The reagents may include either or both the proteins or molecules and the probes required to perform the assay of interest.

In one embodiment, a kit comprises one or more protein chips (*i.e.*, positionally addressable arrays comprising a plurality of different substances, selected from the group consisting of proteins, molecules comprising functional domains of said proteins, whole 30 cells, and protein-containing cellular material, on a solid support, with each different substance being at a different position on the solid support), wherein the plurality of different substances consists of at least 100 different substances per cm^2, and in one or more containers, one or more probes, reagents, or other molecules. The substances of the array can be attached to the surface of wells on the solid support. In another embodiment, the 35 protein chip in the kit can have the protein or probe already attached to the wells of the solid support. In yet another embodiment, the protein chip in the kit can have the reagent(s) or

reaction mixture useful for assaying biological activity of a protein or molecule, or assaying interaction of a probe and a protein or molecule, already attached to the wells of the solid support. In yet another embodiment, the reagent(s) is not attached to the wells of the solid support, but is contained in the wells. In yet another embodiment, the reagent(s) is not

5 attached to the wells of the solid support, but is contained in one or more containers, and can be added to the wells of the solid support. In yet another embodiment, the kit further comprises one or more containers holding a solution reaction mixture for assaying biological activity of a protein or molecule. In yet another embodiment, the kit provides a substrate (e.g., beads) to which probes, proteins or molecules of interest, and/or other

10 reagents useful for carrying out one or more assays, can be attached, after which the substrate with attached probes, proteins, or other reagents can be placed into the wells of the chip.

In another embodiment, one or more protein chips in the kit have, attached to the wells of the solid support, proteins with a biological activity of interest. In another

15 embodiment, one or more protein chips in the kit have, attached to the wells of the solid support, at least 50%, 75%, 90% or 95% of all expressed proteins with the same type of biological activity in the genome of an organism. In a specific embodiment, one or more protein chips in the kit have, attached to the wells of the solid support, at least 50%, 75%, 90% or 95% of all expressed kinases, phosphatases, glycosidase, proteases, acetylases, other

20 group transferring enzymes, nucleic acid binding proteins, hormone-binding proteins or DNA-binding proteins, within the genome of an organism (e.g., of a particular species).

E.   Proteins Useful with the Protein Chips

25 Full-length proteins, portions of full-length proteins, and peptides whether prepared from recombinant overexpression in an organism, produced via fragmentation of larger proteins, or chemically synthesized, are utilized in this invention to form the protein chip. Organisms whose proteins are overexpressed include, but are not limited to, bacteria, yeast, insects, humans, and non-human mammals such as mice, rats, cats, dogs, pigs, cows and

30 horses. Further, fusion proteins in which a defined domain is attached to one of a variety of natural or synthetic proteins can be utilized. Proteins used in this invention can be purified prior to being attached to, or deposited into, the wells of the protein chip, or purified during attachment via the use of reagents which have been previously attached to, or deposited into, the wells of the protein chip. These reagents include those that specifically bind

35 proteins in general, or bind to a particular group of proteins. Proteins can be embedded in artificial or natural membranes (e.g., liposomes, membrane vesicles) prior to, or at the time

of attachment to the protein chip. Alternatively, the proteins can be delivered into the wells of the protein chip.

Proteins used in the protein chips of the present invention are preferably expressed by methods known in the art. The InsectSelect system from Invitrogen (Carlsbad, CA,
5   catalog no. K800-01), a non-lytic, single-vector insect expression system that simplifies expression of high-quality proteins and eliminates the need to generate and amplify virus stocks, is a preferred expression system. The preferred vector in this system is pIB/V5-His TOPO TA vector (catalog no. K890-20). Polymerase chain reaction (PCR) products can be cloned directly into this vector, using the protocols described by the manufacturer, and the
10 proteins are then expressed with N-terminal histidine (His) labels which can be used to purify the expressed protein.

The BAC-TO-BAC™ system, another eukaryotic expression system in insect cells, available from Lifetech (Rockville, MD), is also a preferred expression system. Rather than using homologous recombination, the BAC-TO-BAC™ system generates recombinant
15 baculovirus by relying on site-specific transposition in *E. coli*. Gene expression is driven by the highly active polyhedrin promoter, and therefore can represent up to 25% of the cellular protein in infected insect cells.

VI.   Example I: Analysis of Yeast Protein Kinases Using Protein Chips
20
A. Introduction

The following example exemplifies the various aspects of protein chip production and a method of using the protein chips of the present invention. The protein chip
25 technology of the present invention is suitable for rapidly analyzing large numbers of samples, and therefore this approach was applied to the analysis of nearly all yeast protein kinases. Protein kinases catalyze protein phosphorylation and play a pivotal role in regulating basic cellular functions, such as cell cycle control, signal transduction, DNA replication, gene transcription, protein translation, and energy metabolism[7]. The availability
30 of a complete genome sequence makes it possible to analyze all of the protein kinases encoded by an organism and determine their *in vitro* substrates.

The yeast genome has been sequenced and contains approximately 6200 open reading frames greater than 100 codons in length; 122 of these are predicted to encode protein kinases. Twenty-four of these protein kinase genes have not been studied
35 previously[8]. Except for two histidine protein kinases, all of the yeast protein kinases are

members of the Ser/Thr family; tyrosine kinase family members do not exist although seven protein kinases that phosphorylate serine/threonine and tyrosine have been reported[8].

   With the development of the protein chip technology of the present invention, the high throughput analysis of the biochemical activities of nearly all of the protein kinases
5   from *Saccharomyces cerevisiae* has been conducted as described herein. Protein chips utilized were disposable arrays of 300 nl wells in silicone elastomer sheets placed on top of microscope slides. The high density and small size of the wells allows for high throughput batch processing and simultaneous analysis of many individual samples, requiring only small amounts of protein. Using protein chips of the present invention, *Saccharomyces*
10 *cerevisiae* kinase proteins (119 different kinases in total) were fused to glutathione-S-transferase (GST), overexpressed in yeast, then purified and assayed for their ability to phosphorylate 17 different substrates. Nearly all of the kinases tested (93%) exhibited activities that were at least five-fold higher than controls, on one or more substrates, including 18 of 24 previously uncharacterized kinases. Thirty-two kinases exhibited
15 preferential phosphorylation of one or two substrates. Twenty-seven kinases readily phosphorylated poly(Tyr-Glu). Since only five of these kinases were previously classified as dual function kinases (*i.e.*, they phosphorylate both Ser/Thr and Tyr), these findings greatly expand our knowledge as to which kinases are able to phosphorylate tyrosine residues. Interestingly, these dual specificity kinases often share common amino acid
20 resides that lie near the catalytic region. These results indicate that the protein chip technology of the present invention is useful for high throughput screening of protein biochemical activity, and for the analysis of entire proteomes.

B. Methods

25

1.   Cell Culture, Constructs and Protein Purification

   Using the recombination strategy of Hudson *et al.*[9], 119 of 122 yeast protein kinase genes were cloned into a high copy *URA3* expression vector (pEG(KG)), which produces
30 GST fusion proteins under the control of the galactose-inducible *GAL10* promoter[10]. Briefly, primers complementary to the end of each ORF were purchased from Research Genetics; the ends of these primers contain a common 20 bp sequence. In a second round of PCR, the ends of these products were modified by adding sequences that are homologous to the vector. The PCR products containing the vector sequences at their ends were
35 transformed along with the vector into a *pep4* yeast strain (which lacks several yeast proteases)[10], and Ura[+] colonies were selected. Plasmids were rescued in *E. coli*, verified by

restriction enzyme digestion and the DNA sequence spanning the vector-insert junction was determined using a primer complementary to the vector. For the GST::Cla4 construct, a frame-shift mutation was found in a poly(A) stretch in the amino terminal coding region. Three independent clones were required to find the correct one that maintained reading

5   frame. For five of these genes, two overlapping PCR products were obtained and introduced into yeast cells. Confirmed plasmids were reintroduced into the *pep4* yeast strain for kinase protein purification.

For preparing samples using the 96 well format, 0.75 ml of cells were grown in medium containing raffinose to O.D.(600) about 0.5 in boxes containing 2 ml wells; two

10   wells were used for each strain. Galactose was added to a final concentration of 4% to induce protein expression, and the cells were incubated for 4 hrs. The cultures of the same strain were combined, washed once with 500 μl of lysis buffer, resuspended in 200 μl of lysis buffer, and transferred into a 96 X 0.5 ml plate (Dot Scientific, USA) containing 100 μl chilled glass beads. Cells were lysed in the box by repeated vortexing at 4 °C and the

15   GST fusion proteins were purified from these strains using glutathione beads and standard protocols[20] in a 96 well format. The purity of five purified GST::kinase proteins (Swe1, Ptk2, Pkh1, Hog1, Pbs2) was determined by comparing the Coomasie staining patterns of the purified proteins with the patterns obtained by immunoblot analysis using anti-GST antibodies. The results indicated that the purified proteins were more than 90% pure. To

20   purify the activated form of Hog1, the cells were challenged with 0.4 M NaCl in the last five minutes of the induction. Protein kinase activity was stable for at least two months at –70°C with little or no loss of kinase activity.

2.   Chip Fabrication and Protein Attachment

25

Chips were made from the silicone elastomer, polydimethylsiloxane (PDMS) (Dow Chemical, USA), which was cast over microfabricated molds. Liquid PDMS was poured over the molds and, after curing (at least 4 hours at 65 °C), flexible silicone elastomer array sheets were then peeled from the reusable molds. Although PDMS can be readily cast over

30   microlithographically fabricated structures, for the purposes of the kinase assay described herein, molds made from sheets of acrylic patterned with a computer-controlled laser milling tool (Universal Laser Systems, USA) sufficed.

Over 30 different arrays were tested. The variables tested were width and depth of the wells (widths ranging from 100 μm to 2.5 mm, depths from 100 μm to 1 mm), spacing

35   between wells (100 μm to 1 mm), configuration (either rectangular arrays or closest packed), and well shape (square versus round). The use of laser milled acrylic molds

offered a fast and inexpensive method to realize a large number prototype molds of varying parameters.

To determine the conditions that maximize protein attachment to the wells, PDMS was treated with either 5 M H_2SO_4, 10 M NaOH, hydrogen peroxide or a 3-

5   glycidooxypropyltrimethoxysilane linker (GPTS)(Aldrich, USA) [11,12]. GPTS treatment resulted in the greatest adsorption of protein to the wells relative to untreated PDMS or PDMS treated other ways. Briefly, after washing with 100% EtOH three times at room temperature, the chips were immersed in 1% GPTS solution (95% EtOH, 16 mM HOAc) with shaking for 1 hr at room temperature. After three washes with 95% EtOH, the chips

10 were cured at 135°C for 2 hrs under vacuum. Cured chips can be stored in dry Argon for months[12]. To attach proteins to the chips, protein solutions were added to the wells and incubated on ice for 1 to 2 hours. After rinsing with cold HEPES buffer (10 mM HEPES, 100 mM NaCl, pH 7.0) three times, the wells were blocked with 1% BSA in PBS (Sigma, USA) on ice for > 1 hr. Because of the use of GPTS, any reagent containing primary amine

15 groups was avoided.

To determine the concentration of proteins that can be linked to the treated PDMS, horseradish peroxidase (HRP) anti-mouse Ig (Amersham, USA) was attached to the chip using serial dilutions of the enzyme. After extensive washing with PBS, the bound antibodies were detected using an enhanced chemiluminescent (ECL) detection method

20 (Amersham, USA). Up to 8×10^{-9} µg/µm^2 of protein can be attached to the surface; a minimum 8×10^{-13} µg/µm^2 is required for detection by our immunostaining methods.

3.   Immunoblotting, Kinase Assay and Data Acquisition

25   Immunoblot analysis was performed as described[34]. GST::protein kinases were tested for *in vitro* kinase activity[13] using ^{33}P–γ-ATP. In the autophosphorylation assay, the GST::kinases were directly adhered to GPTS-treated PDMS and the *in vitro* reactions carried out with ^{33}P-γ-ATP in appropriate buffer. In the substrate reactions, the substrate was adhered to the wells via GPTS, and the wells were washed with HEPES buffer and

30   blocked with 1% BSA, before kinase, ^{33}P-γ-ATP and buffer were added. The total reaction volume was kept below 0.5 µl per reaction. After incubation for 30 minutes at 30°C, the chips were washed extensively, and exposed to both X-ray film and a Molecular Dynamics PhosphorImager, which has a resolution of 50 µm and is quantitative. For twelve substrates each kinase assay was repeated at least twice; for the remaining five, the assays were

35   performed once.

To determine substrate specificity, specificity index (SI) was calculated using the following formula: $SI_{ir} = F_{ir} / [(F_{i1} + F_{i2} + \bullet \bullet \bullet \bullet \bullet + F_{ir}) / r]$, where i represents the ID of kinase used, r represents the ID of a substrate, and F_{ir} represents the fold increase of a kinase i on substrate r compared with GST alone.

4.   Kinase Sequence Alignments and Phylogenetic Trees

Multiple sequence alignments based on the core kinase catalytic domain subsequences of the 107 protein kinases were generated with the CLUSTAL W algorithm[35], using the Gonnet 250 scoring matrix[36]. Kinase catalytic domain sequences were obtained from the SWISS-PROT[37], PIR[38], and GenBank[39] databases. For those kinases whose catalytic domains are not yet annotated (DBF4/YDR052C and SLN1/YIL147C), probable kinase subsequences were inferred from alignments with other kinase subsequences in the data set with the FASTA algorithm[40,41] using the BLOSUM 50 scoring matrix[42]. Protein subsequences corresponding to the eleven core catalytic subdomains[43] were extracted from the alignments, and the phylogenetic trees were computed with the PROTPARS[44] program (Figure 5a).

5.   Functional Grouping of Protein Chip Data

To visualize the approximate functional relationships between protein kinases relative to the experimental data, kinases were hierarchically ordered based on their ability to phosphorylate the 12 different substrates (data available on web site http://bioinfo.mbb.yale.edu/genome/yeast/chip as of August 17, 2000). A profile corresponding to the -/+ activity of the 107 protein kinases to each of the substrates was recorded, with discretized values in [0,1]. Matrices were derived from the pairwise Hamming distances between experimental profiles, and unrooted phylogenies were computed using the Fitch-Margoliash least-squares estimation method[45] as implemented in the FITCH program34 of the PHYLIP software package[44]. In each case, the input order of taxa was randomized to negate any inherent bias in the organization of the data set, and optimal hierarchies were obtained through global rearrangements of the tree structures.

C. Results

1.   Yeast Kinase Cloning and Protein Purification

Using a recombination-directed cloning strategy[9], we attempted to clone the entire coding regions of 122 yeast protein kinase genes in a high copy expression vector (pEG(KG)) that produces GST fusion proteins under the control of the galactose-inducible *GAL10* promoter[10] (Figure 1a). GST::kinase constructs were rescued into *E. coli*, and

5 sequences at the 5'-end of each construct were determined. Using this strategy, 119 of the 122 yeast protein kinase genes were cloned in-frame. The three kinase genes that were not cloned are very large (4.5-8.3 kb).

The GST:kinase fusion proteins were overproduced in yeast and purified from 50 ml cultures using glutathione beads and standard protocols[11]. For the case of Hog1 the yeast

10 cells were treated with high salt to activate the enzyme in the last five minutes of induction; for the rest of the kinases, synthetic media (URA'/raffinose) was used. Immunoblot analysis of all 119 fusions using anti-GST antibodies revealed that 105 of the yeast strains produced detectable GST::fusion proteins; in most cases the fusions were full length. Up to 1 μg of fusion protein per ml of starting culture was obtained (Figure 1b). However, 14 of

15 119 GST::kinase samples were not detected by immunoblotting analysis. Presumably, these proteins are not stably overproduced in the *pep4* protease-deficient strain used, or these proteins may form insoluble aggregates that do not purify using our procedures. Although this procedure was successful, purification of GST fusion proteins using 50 ml cultures is a time-consuming process and not applicable for preparing thousands of samples.

20 Therefore, a procedure for growing cells in a 96 well format was developed (see Methods). Using this procedure, 119 GST fusions were prepared and purified in six hours with about two-fold higher yields per ml of starting culture relative to the 50 ml method.

2.   Protein Chip Design

25

Protein chips were developed to conduct high throughput biochemical assays of 119 yeast protein kinases (Figure 2). These chips consist of an array of wells in a disposable silicone elastomer polydimethylsiloxane (PDMS)[11]. Arrays of wells allow small volumes of different probes to be densely packed on a single chip yet remain physically segregated

30 during subsequent batch processing. Proteins were covalently attached to the wells using a linker 3-glycidooxypropyltrimethoxysilane (GPTS)[12]. Up to 8×10^{-9} μg/μm² of protein can be attached to the surface (see Methods).

For the purposes of the protein kinase assays, the protein chip technology was configured to be compatible with standard sample handling and recording equipment.

35 Using radioisotope labeling (^{33}P), the kinase assays described below, and manual loading, a variety of array configurations were tested. The following chips produced the best results:

- 32 -

round wells with 1.4 mm diameter and 300 μm deep (approximately 300 nl), in a 10x14 rectangular array configuration with a 1.8 mm pitch. A master mold of twelve of these arrays were produced, and a large number of arrays were repeatedly cast for the protein kinase analysis. Chips were placed atop microscope slides for handling purposes (Figure 2a); the arrays covered slightly more than one third of a standard microscope slide and two arrays per slide were typically used (Figure 2b). Although a manual pipette method to place proteins in each well was employed, automated techniques may also be used. In addition, this protein chip configuration may also be used with other labeling methods, such as by using fluorescently labeled antibodies to phosphoproteins, and subsequent detection of immunofluorescence.

3. Large-scale Kinase Assays Using Protein Chips

All 119 GST::protein kinases were tested for *in vitro* kinase activity[13] in 17 different assays using ^{33}P–γ-ATP and 17 different chips. Each chip was assayed using a different substrate, as follows: 1) Autophosphorylation, 2) Bovine Histone H1 (a common kinase substrate), 3) Bovine Casein (a common substrate), 4) Myelin basic protein (a common substrate), 5) Axl2 C terminus-GST (Axl2 is a transmembrane phosphoprotein involved in budding)[14], 6) Rad9 (a phosphoprotein involved in the DNA damage checkpoint)[15], 7) Gic2 (a phosphoprotein involved in budding)[16], 8) Red1 (a meiotic phosphoprotein important for chromosome synapsis)[17], 9) Mek1 (a meiotic protein kinase important for chromosome synapsis)[18], 10) Poly(tyrosine-glutamate 1:4) (poly(Tyr-Glu)); a tyrosine kinase substrate)[19], 11) Ptk2 (a small molecule transport protein)[20], 12) Hsl1 (a protein kinase involved in cell cycle regulation)[21], 13) Swi6 (a phosphotranscription factor involved in G1/S control)[22], 14) Tub4 (a protein involved in microtubule nucleation)[23], 15) Hog1 (a protein kinase involved in osmoregulation)[24], 16) Hog1 (an inactive form of the kinase), and 17) GST (a control). For the autophosphorylation assay, the kinases were directly adhered to the treated PDMS wells and ^{33}P-γ-ATP was added; for substrate reactions, the substrates were bound to the wells, and then kinases and ^{33}P-γ-ATP were added. After the reactions were completed, the slides were washed and the phosphorylation signals were acquired and quantified using a high resolution phosphoimager. Examples are shown in Figure 3. To identify kinase activities, the quantified signals were converted into fold increases relative to GST controls and plotted for further analysis (Figure 4a).

As shown in Figure 4a, most (93.3%) kinases exhibited activity five-fold or greater over background for at least one substrate. As expected, Hrr25, Pbs2 and Mek1 phosphorylated their known substrates[25-27], Swi6 (400-fold higher than the GST control),

Hog1 (10-fold higher) and Red1 (10-fold higher), respectively. The results of this assay demonstrated that 18 of the 24 predicted protein kinases have not been studied previously phosphorylate one or more substrates, as do several unconventional kinases[8], including the histidine kinases (Sln1, Yil042c) and phospholipid kinases (e.g Mec1).

5   To determine substrate specificity, the activity of a particular kinase was further normalized against the average of its activity against all substrates. Several examples are shown in Figure 4b; all the data are available at http://bioinfo.mbb.yale.edu/genome/yeast/chip. Thirty-two kinases exhibited substrate specificity on a particular substrate with specificity index (SI; see Methods) equal or higher

10   than 2, and reciprocally, most substrates are preferentially phosphorylated by a particular protein kinase or set of kinases. For example, the C terminus of Axl2, a protein involved in yeast cell budding, is preferentially phosphorylated by Dbf20, Kin2, Yak1 and Ste20 relative to other protein. Interestingly, previous studies found that Ste20 was localized at the tip of emerging buds similar to Axl2, and a *ste20Δ/cla4ᵗˢ* mutant is unable to bud or

15   form fully polarized actin patches or cables[28]. Another example is the phosphoprotein Gic2, which is also involved in budding[16]. Ste20 and Skm1 strongly phosphorylate Gic2 (Figure 4b). Previous studies suggested that Cdc42 interacts with Gic2, Cla4[29], Ste20 and Skm1. Our results raise the possibility that Cdc42 may function to promote the phosphorylation of Gic2 by recruiting Ste20 and/or Skm1.

20

4.   Yeast Contain Many Dual Specific Kinases

Of particular interest are the dual specificity kinases, *i.e.*, those enzymes that phosphorylate both Ser/Thr and tyrosine. Based on sequence analysis, all but two yeast

25   protein kinases belong to the Ser/Thr family of protein kinases; however, at the time of the study, seven protein kinases (Mps1, Rad53, Swe1, Ime2, Ste7, Hrr25, and Mck1) were reported to be dual specificity kinases[19]. We confirmed that Swe1, Mps1, Ime2, and Hrr25 readily phosphorylate poly(Tyr-Glu), but we did not detect any tyrosine kinase activity for Ste7, Rad53 or Mck1. Mck1 did not show strong activity in any of our assays; however,

30   Ste7 and Rad53 are very active in other assays. Thus, their inability to phosphorylate poly(Tyr-Glu) indicates that they are either very weak tyrosine kinases in general or at least are weak with the poly(Tyr-Glu) substrate. Consistent with the latter possibility, others have found that poly(Tyr-Glu) is a very poor substrate for Rad53 (Ref 19; D. Stern, pers. comm.). Interestingly, we found that 23 other kinases also efficiently use poly(Tyr-Glu) as

35   a substrate, indicating that there are at least 27 kinases in yeast that are capable of acting *in vitro* as dual specificity kinases. One of these, Rim1, was recently shown to phosphorylate

a Tyr residue on its in vivo substrate, Ime2, indicating that it is a *bona fide* dual specificity kinase[30]. In summary, this experiment roughly tripled the number of kinases capable of acting as dual specificity kinases, and has raised questions about some of those classified as such kinases.

5

5.   Correlation Between Functional Specificity and Amino Sequences of the Poly(Tyr-Glu) Kinases

The large-scale analysis of yeast protein kinases allows us to compare the functional
10   relationship of the protein kinases to one another. We found that many of the kinases that phosphorylate poly(Tyr-Glu) are related to one another in their amino acid sequences: 70% of the poly(Tyr-Glu) kinases cluster into distinct four groups on a dendrogram in which the kinases are organized relative to one another based on sequence similarity of their conserved protein kinase domains (Figure 5a). Further examination of the amino acid
15   sequence reveals four types of amino acids that are preferentially found in the poly(Tyr-Glu) class of kinases relative to the kinases that do not use poly(Tyr-Glu) as a substrate (three are lysines and one is a methionine); one residue (an asparagine) was preferentially located in the kinases that do not readily use poly(Tyr-Glu) as a substrate (Figure 5b). Most of the residues lie near the catalytic portion of the molecule (Figure 5b)[31], suggesting that
20   they may play a role in substrate recognition.

D.   Discussion

1.   Large-scale Analysis of Protein Kinases

25

This study employed a novel protein chip technology to characterize the activities of 119 protein kinases for 17 different substrates. We found that particular proteins are preferred substrates for particular protein kinases, and vice versa, many protein kinases prefer particular substrates. One concern with these studies is that it is possible that kinases
30   other than the desired enzyme are contaminating our preparations. Although this cannot be rigorously ruled out, analysis of five of our samples by Coomasie staining and immunoblot staining with anti-GST does not reveal any detectable bands in our preparation that are not GST fusions (see methods).

It is important to note that *in vitro* assays do not ensure that a substrate for a
35   particular kinase *in vitro* is phosphorylated by the same kinase *in vivo*. Instead, these experiments indicate that certain proteins are capable of serving as substrates for specific

kinases, thereby allowing further analysis. In this respect, these assays are analogous to two-hybrid studies in which candidate interactions are detected. Further experimentation is necessary to determine if the processes normally occur *in vivo*.

Consistent with the idea that many of the substrates are likely to be bonafide substrates *in vivo* is the observation that three kinases, Hrr25, Pbs2 and Mek1, phosphorylate their known substrates in our assays. Furthermore, many of the kinases (*e.g.*,Ste20) co-localize with their *in vitro* substrates (*e.g.*,Axl2). Thus, we expect many of the kinases that phosphorylate substrates in our *in vitro* assays are likely also to do so *in vivo*.

Although most of the kinases were active in our assays, several were not. Presumably, our preparations of these latter kinases either lack sufficient quantities of an activator or were not purified under activating conditions. For example, Cdc28 which was not active in our assays, might be lacking its activating cyclins. For the case of Hog1, cells were treated with high salt to activate the enzyme. Since nearly all of our kinase preparations did exhibit activity, we presume that at least some of the enzyme in the preparation has been properly activated and/or contains the necessary cofactors. It is likely that the overexpression of these enzymes in their native organism contributes significantly to the high success of obtaining active enzymes.

Using the assays on the protein chip, many kinases that utilize poly(Tyr-Glu) were identified. The large-scale analysis of many kinases allowed the novel approach of correlating functional specificity of poly(Tyr-Glu) kinases with specific amino acid sequences. Many of the residues of the kinases that phosphorylate poly(Tyr-Glu) contain basic residues. This might be expected if there were electrostatic interactions between the kinases residues and the Glu residues. However, the roles of some of the other residues are not obvious such as the Met residues on the kinases that phosphorylate poly(Tyr-Glu) and the Asn on those that do not. These kinase residues may confer substrate specificity by other mechanisms. Regardless, analysis of additional substrates should allow further correlation of functional specificity with protein kinase sequence for all protein kinases.

2.   Protein chip technology

In addition to the rapid analysis of large number of samples, the protein chip technology described here has significant advantages over conventional methods. 1) The chip-based assays have high signal-to-noise ratios. We found that the signal-to-noise ratio exhibited using the protein chips is much better (>10 fold) than that observed for traditional microtiter dish assays (data not shown). Presumably this is due to the fact that ^{33}P-γ-ATP

does not bind the PDMS as much as microtiter dishes. 2) The amount of material needed is very small. Reactions volumes are 1/20-1/40 the amount used in the 384-well microtiter dishes; less than 20 ng of protein kinase was used in each reaction. 3) The enzymatic assays using protein chips are extremely sensitive. Even though only 105 fusions were detectable by immunoblot analysis, 112 exhibited enzymatic activity greater than five-fold over background for at least one substrate. For example, Mps1 consistently exhibits the strongest activity in many of the kinase assays even though we have not been able to detect this fusion protein by immunoblot analysis (see Figures 1b and 3a). 4) Finally, the chips are inexpensive; the material costs less than eight cents for each array. The microfabricated molds are also easy to make and inexpensive.

In addition to the analysis of protein kinases, this protein chip technology is also applicable to a wide variety of additional assays, such as ATP and GTP binding assays, nuclease assays, helicase assays and protein-protein interaction assays. Recently, in an independent study, Phizicky and coworkers expressed yeast proteins as GST fusions under the much weaker *CUP1* promotor[6]. Although the quality of their clones has not been established, they were able to identify biochemical activities using pools of yeast strains containing the fusion proteins. The advantage of our protein chip approach is that all samples can be analyzed in a single experiment. Furthermore, although this study used wells which have the advantage of segregating samples, flat PDMS chips and glass slides can also be used for different assays; these have the advantage that they can be used with standard pinning tool microarrayers. This technology can also be applied to facilitate high-throughput drug screening in which one can screen for compounds that inhibit or activate enzymatic activities of any gene products of interest. Since these assays will be carried out at the protein level, the results will be more direct and meaningful to the molecular function of the protein.

We configured the protein chip technology for a specific protein kinase assay using commonly available sample handling and recording equipment. For this purpose, array dimensions remained relatively large compared to dimensions readily available with microfabricated silicone elastomer structures[32]. We have cast PDMS structures with feature sizes two orders of magnitude smaller than those reported here using microlithographically fabricated molds, while others have reported submicron feature sizes in microfabricated structures[33]. These results indicate that well densities of microfabricated protein chips can be readily increased by several orders of magnitude. The protein chip technology reported here is readily scalable.

In conclusion, an inexpensive, disposable protein chip technology was developed for high throughput screening of protein biochemical activity. Utility was demonstrated

through the analysis of 119 protein kinases from *Saccharomyces cerevisiae* assayed for phosphorylation of 17 different substrates. These protein chips permit the simultaneous measurement of hundreds of protein samples. The use of microfabricated arrays of wells as the basis of the chip technology allows array densities to be readily increased by several

5   orders of magnitude. With the development of appropriate sample handling and measurement techniques, these protein chips can be adapted for the simultaneous assay of several thousand to millions of samples.

E. References

10

1.   Fields, S., Kohara, Y., and Lockhart, D.J. Functional genomics. *Proc. Natl. Acad. Sci.* **96**, 8825-26 (1999).

2.   Goffeau, A., *et al*. Life with 6000 genes. *Science* **274**, 563-567 (1996).

3.   DeRisi, J.L., Iyer, V.R., and Brown, P.O. Exploring the metabolic and genetic

15   control of gene expression on a genomic scale. *Science* **278**, 680-686 (1997).

4.   Winzeler, E.A., *et al*. Functional characterization of the *S. cerevisiae* genome by gene deletion and parallel analysis. *Science* **285**, 901-906 (1999).

5.   Heyman, J.A., *et al*. Genome-scale cloning and expression of individual open reading frames using topoisomerase I-mediated ligation. *Genome Res.* **9**, 383-392

20   (1999).

6.   Martzen, M.R., *et al*. A biochemical genomics approach for identifying genes by the activity of their products. *Science* **286**, 1153-1155 (1999).

7.   Plowman G.D., Sudarsanam S., Bingham J., Whyte D., and Hunter T. The protein kinases of *Caenorhabditis elegans*: A model for signal transduction in multicellular

25   organisms. *Proc. Natl. Acad. Sci.* **96**, 13603-12610 (1999).

8.   Hunter, T., & Plowman, G.D. The protein kinases of budding yeast: six score and more. *TIBS* **22**, 18-22 (1997).

9.   Hudson, J.R., *et al*. The complete set of predicted genes from *Saccharomyces cerevisiae* in a readily usable form. *Genome Res.* **7**, 1169-1173 (1997).

30 10.   Mitchell, D.A., Marshall, T.K., and Deschenes, R.J. Vector for the inducible overexpression of glutathione S-transferase fusion protein in yeast. *Yeast* **9**, 715-23 (1993).

11.   Rogers, Y.-H., *et al*. Immobilization of oligonucleotides onto a glass support via disulfide bonds: a method for preparation of DNA microarrays. *Analy. Biochem.*

35   **266**, 23-30 (1999).

12. Stimpson, D.J., *et al.* Real-time detection of DNA hybridization and melting on oligonucleotide arrays by using optical wave guides. *Proc. Natl. Acad. Sci.* **92**, 6379-6383 (1995).

13. Hunter, T. & Sefton, B.M. Protein phosphorylation. *Meth. in Enzym.* **200**, 35-83 (1991).

14. Roemer, T.K., *et al.* Selection of axial growth sites in yeast requires Axl2p, a novel plasma membrane glycoprotein. *Genes & Dev.* **10**, 777-793 (1996).

15. Weinert, T.A. & Hartwell, L.H. Cell cycle arrest of cdc mutants and specificity of the RAD9 checkpoint. *Genetics* **134**, 63-80 (1993).

16. Jaquenoud, M., Gulli, M.P., Peter, K., and Peter, M. The Cdc42p effector Gic2p is targeted for ubiquitin-dependent degradation by the SCFGrr1 complex. *EMBO J.* **17**, 5360-5373 (1998).

17. Menees, T.M., Ross-MacDonald, P.B., and Roeder, G.S. MEI4, a meiosis-specific yeast gene required for chromosome synapsis. *Mol. Cell Biol.* **12**, 1340-1351 (1992).

18. Bailis, J.M., & Roeder, G.S. Synaptonemal complex morphogenesis and sister-chromatid cohesion require Mek1-dependent phosphorylation of a meiotic chromosomal protein. *Genes & Dev.* **12**, 3551-3563 (1998).

19. Stern, D.F., Zheng, P., Beidler, D.R., and Zerillo, C. Spk1, a new kinase from *Saccharomyces cerevisiae* phosphorylates proteins on serine, threonine, and tyrosine. *Mol. Cell. Biol.* **11**, 987-1001 (1991).

20. Kaouass, M., *et al.* The STK2 gene, which encodes a putative Ser/Thr protein kinase, is required for high-affinity spermidine transport in *Saccharomyces cerevisiae*. *Mol. Cell Biol.* **17**, 2994-3004 (1997).

21. Barral, Y., Parra, M., Bidlingmaier, S., and Snyder, M. Nim1-related kinases coordinate cell cycle progression with the organization of the peripheral cytoskeleton in yeast. *Genes & Dev.* **13**, 176-187 (1999).

22. Madden, K., Sheu, Y.-J., Baetz, K., Andrews, B., and Snyder, M. SBF cell cycle regulator as a target of the yeast PKC-MAP kinase pathway. *Science* **275**, 1781-1784 (1997).

23. Sobel, S.G. & Snyder, M. A highly divergent gamma-tubulin gene is essential for cell growth and proper microtubule organization in *Saccharomyces cerevisiae*. *J. Cell Biol.* **131**, 1775-1788 (1995).

24. Ferrigno, P., Posas, F., Koepp, D., Saito, H., and Silver, P.A. Regulated nucleo/cytoplasmic exchange of HOG1 MAPK requires the importin beta homologs NMD5 and XPO1. *EMBO J.* **17**, 5606-5614 (1998).

25.   Ho, U., Mason, S., Kobayashi, R., Heokstra, M., and Andrew, B. Role of the casein kinase I isoform, Hrr25, and the cell cycle-regulatory transcription factor, SBF, in the transcriptional response to DNA damage in *Saccharomyces cerevisiae*. *Proc. Natl. Acad. Sci.* **94**, 581-586 (1997).

5 26.   Wurgler-Murphy, S.M., Maeda, T., Witten, E.A., and Saito, H. Regulation of the *Saccharomyces cerevisiae* HOG1 mitogen-activated protein kinase by the PTP2 and PTP3 protein tyrosine phosphatases. *Mol. Cell Biol.* **17**, 1289-1297 (1997).

27.   Santos, T. & Hollingsworth, N.M. Red1p, a MEK1-dependent phosphoprotein that physically interacts with Hop1p during meiosis in yeast. *J. Biol. Chem.* **274**, 1783-

10   1790 (1999).

28.   Holly, S.P. & Blumer, K.J. PAK-family kinases regulate cell and actin polarization throughout the cell cycle of *Saccharomyces cerevisiae*. *J. Cell Biol.* **147**, 845-856 (1999).

29.   Richman, T.J., Sawyer, M.M., and Johnson, D.I. The Cdc42p GTPase is involved in

15   a G2/M morphogenetic checkpoint regulating the apical-isotropic switch and nuclear division in yeast. *J. Biol. Chem.* **274**, 16861-16870 (1999).

30.   Malathi, K., Xiao, Y., and Mitchell, A.P. Catalytic roles of yeast GSK3beta/shaggy homolog Rim11p in meiotic activation. *Genetics* **153**, 1145-1152 (1999).

31.   Owen, D.J., Noble, M.E., Garman, E.F., Papageorgiou, A.C., and Johnson, L.N.

20   Two structures of the catalytic domain of phosphorylase kinase: an active protein kinase complexed with substrate analogue and product. *Structure*, **3**, 467-474 (1995).

32.   Xia, Y. & Whitesides, G.M. *Angew. Chem. Int.* Ed. **37**, 550- (1997).

33.   Jackman, R.J., Duffy, D.C., Cherniavskaya, O., and Whitesides, G.M. Using

25   elastomeric membranes as dry resists and for dry lift-off. *Langmuir*, **15**, 2973-2984 (1999).

34.   Mylin, L.M., Hofmann, K.J., Schultz, L.D., and Hopper, J.E. Regulated GAL4 expression cassette providing controllable and high-level output from high-copy galactose promoters in yeast. *Methods Enzymol.* **185**, 297-308 (1990).

30 35. Higgins, D.G., Thompson, J.D., and Gibson, T.J. Using CLUSTAL for multiple sequence alignments. *Methods Enzymol.* **266**, 383-402 (1996).

36.   Gonnet, G.H., Cohen, M.A., and Benner, S.A. Exhaustive matching of the entire protein sequence database. *Science.* **256**, 1443-1445 (1992).

37.   Bairoch, A. & Apweiler, R. The SWISS-PROT protein sequence data bank and its

35   supplement TrEMBL. *Nucleic Acids Res.* **27**, 49-54 (1999).

38. Barker, W.C., *et al.* The PIR-International Protein Sequence Database. *Nucleic Acids Res.* **27**(1), 39-43 (1999).

39. Benson, D.A., *et al.* GenBank. *Nucleic Acids Res.* **27**, 12-17 (1999).

40. Lipman, D.J. & Pearson, W.R. Rapid and sensitive protein similarity searches. *Science.* **277**, 1435-1441 (1985).

41. Pearson, W.R. & Lipman, D.J. Improved tools for biological sequence comparison. *Proc. Natl. Acad. Sci.* **85**, 2444-2448 (1988).

42. Dayhoff, M.O., Schwartz, R.M., and Orcutt, B.C. A model of evolutionary change in proteins. In *Atlas of Protein Sequence and Structure*, M.O. Dayhoff, Ed. Washington, DC: National Biomedical Research Foundation. pp 345-352 (1978).

43. Hanks, S.K. & Hunter, T. Protein Kinases 6. The eukaryotic protein kinase superfamily: kinase (catalytic) domain structure and classification. *FASEB J.* **9**, 576-596 (1995).

44. Felsenstein, J. PHYLIP - Phylogeny Inference Package (Version 3.2). *Cladistics.* **5**, 164-166 (1989).

45. Fitch, W.M. & Margoliash, E. Construction of phylogenetic trees. *Science.* **155**, 279-284 (1967).

VII.   Example II: Analysis of Yeast Protein Kinase Activity Using Protein Chips

A.   Introduction

The following example presents three protocols that, for illustration purposes only, provide different methods of using the protein chips of the present invention to assay for protein kinase activity.

1.   Assay Methods for Protein Kinase Activity

i.   Autophosphorylation Activity

(1)   Protein chips were washed three times with 100% EtOH at room temperature. The chips were then coated with the linker GPTS (1% in 95% EtOH) at room temperature for one hour with shaking. After washing with 100% EtOH three times, the chips were dried at 130°C for 1.5 hours under vacuum.

(2)   GST::yeast protein kinases, one kinase species per well, were bound to the wells of the protein chip by incubation for at least one hour. The chip was further blocked by 1% BSA.

(3)   Kinase buffer and a ^{33}P-γ-ATP probe was added to each well, and incubated at 30°C for 30 minutes. The chip was washed extensively after the phosphorylation reaction was completed.

(4)   The specific ^{33}P-γ-ATP signal, representing autophosphorylation, was
5   detected and quantified by a phosphoimager.

ii.   Kinase Activity - Protocol I

(1)   Protein chips were washed three times with 100% EtOH at room temperature. The chips were then coated with the linker GPTS (1% in 95% EtOH) at room
10   temperature for one hour with shaking. After washing with 100% EtOH three times, the chips were dried at 130°C for 1.5 hours under vacuum.

(2)   A substrate (for example, GST::yeast protein) was bound to the chips by incubation for one or more hours. The chip was further blocked by 1% BSA, and the chip was washed.

15   (3)   A different protein kinase was added to each well of the protein chip, along with kinase buffer and ^{33}P-γ-ATP, and incubated at 30°C for 30 minutes. The protein chip was washed extensively after the phosphorylation reaction was completed.

(4)   The specific ^{33}P-γ-ATP signal, representing phosphorylation of the substrate protein by the protein kinase probe, was detected and quantified by a
20   phosphoimager.

iii.   Kinase activity - Protocol II

(1)   Protein chips were washed three times with 100% EtOH at room temperature. The chips were then coated with the linker GPTS (1% in 95% EtOH) at room
25   temperature for one hour with shaking. After washing with 100% EtOH three times, the chips were dried at 130°C for 1.5 hours under vacuum.

(2)   A substrate (for example, GST::yeast protein) was bound to the chips by incubation for one or more hours. The chip was further blocked by 1% BSA and the chip was washed.

30   (3)   A different protein kinase was added to each well of the protein chip, along with kinase buffer and P-γ-ATP, and incubated at 30°C for 30 minutes. The protein chip was washed extensively after the phosphorylation reaction was completed. The chip was incubated with iodoacetyl-LC-biotin in the dark at room temperature overnight.

(4)   After washing, the chip was probed with fluorescent-labeled avidin to
35   detect the phosphorylation signals.

(5)   The chip was then scanned using an Axon Genepix 4000A scanner, which was modified with a lens having an increased depth of focus of about 300-400 microns. The modifications allow scanning of surfaces mounted on a slide (*e.g.*, the PDMS microarrays of the present invention), which would otherwise be out of the plane of focus.

5   Using the modified Axon Genepix 4000A scanner, the arrays were scanned to acquire and quantify fluorescent signals.

VIII.   Example III: Analysis of Protein-Protein Interactions Using Protein Chips

10

A protein of interest ("probe protein") is recombinantly expressed in and purified from *E. coli* as a labeled fusion protein using standard protocols. The target proteins are attached to the wells of the chip, with a different target protein in each well. The purified probe protein is introduced into each well of the chip, and incubated for several hours or

15   more. The chip is washed and probed with either: a) antibodies to the probe protein, or b) antibodies to the label on the fusion protein. The antibodies are labeled with a fluorescent label, such as Cy3 or Cy5, or are detected using a fluorescently labeled secondary antibody that detects the first antibody.

The following examples provide, for illustration purposes only, methods of using

20   the protein chips of the present invention to assay for proteases, nucleases, or G-protein receptors. Protein-protein interactions generally can be assayed using the following or a similar method.

A.   Analysis of Protease Activity

25

Protease activity is assayed in the following way. First, protein probes are prepared consisting of various combinations of amino acids, with a C-terminal or N-terminal mass spectroscopic label attached, with the only proviso being that the molecular weight of the label should be sufficiently large so that all labeled cleavage products of the protein can be

30   detected. The protein probe is contacted with proteases attached to a protein chip at 37°C. After incubation at 37°C for an appropriate period of time, and washing with acetonitrile and trifluoroacetic acid, protease activity is measured by detecting the proteolytic products using mass spectrometry. This assay provides information regarding both the proteolytic activity and specificity of the proteases attached to the protein chip.

35   Another rapid assay for protease activity analysis is to attach proteins of known sequence to the chip. The substrate proteins are fluorescently labeled at the end not

- 43 -

attached to the chip. Upon incubation with the protease(s) of interest, the fluorescent label is lost upon proteolysis, such that decreases in fluorescence indicate the presence and extent of protease activity. This same type of assay can be carried out wherein the protein substrates are attached to beads placed in the wells of the chips.

5

B.   Analysis of Nuclease Activity

Nuclease activity is assessed in the same manner as described for protease activity, above, except that nucleic acid probes/substrates are substituted for protein
10 probes/substrates. As such, fluorescently tagged nucleic acid fragments that are released by nuclease activity can be detected by fluorescence, or the nucleic acid fragments can be detected directly by mass spectrometry.

C.   Analysis of G-Protein Coupled Receptors

15

In another type of assay, compounds that bind G-protein coupled receptors are identified. Initially, the G-protein receptor is cloned as a GST fusion protein, with the GST portion attached to the C terminus of the G-protein because the C-terminus is generally not involved with determining probe specificity. The G-protein::GST fusion proteins are
20 attached to the wells, preferably by association with glutathione. The G-protein receptors are then incubated with a mixture of compounds, such as a combinatorial chemical library or a peptide library. After washing, bound probes are eluted, for example by the addition of 25% acetonitrile/0.05% trichloroacetic acid. The eluted material is then be loaded into a MALDI mass spectrometer and the nature of the bound probes identified.

25

IX.   Example IV: Analysis of Protein Kinases Inhibition by Specific Inhibitors Using Protein Chips

30   The following description provides, for exemplary purposes only, methods of using the protein chips of the present invention to examine protein kinases for sensitivity to protein kinase inhibitors. Protein-protein interactions generally can be assayed using the following or similar method.

Substrates were bound to the surface of the GPTS-treated microwells on the protein
35 chip at room temperature for one hour, then blocked with 1% BSA and 100 mM Tris pH 7.5, and washed three times with TBS buffer. Kinases and different concentrations of

kinase inhibitors were added to the microwells in the presence of $^{33}P\gamma$-ATP. The phosphorylation reaction was carried out at 30°C for thirty minutes. After completion of the reaction, the protein chip was washed extensively with TBS buffer at room temperature, and then allowed to dry. Phosphorylation signals were obtained by exposing the protein chip to either X-ray film or a phosphoimager.

A human protein kinase A (PKA), a human map kinase (MAPK), three yeast PKA homologs (TPK1, TPK2 and TPK3), and two other yeast protein kinases (HSL1 and RCK1) were tested against two substrates (*i.e.*, a protein substrate for PKA and a commonly used kinase substrate, MBP) using different concentrations of PKIα (a specific human PKA inhibitor) or SB202190 (a MAPK inhibitor). As shown in Figure 7, PKIα specifically inhibited PKA activities on both peptide and MBP substrates. However, PKIα did not inhibit the three yeast PKA homologs (TPK1, TPK2, TPK3) or the other two yeast protein kinases tested, HSL1 and RCK1). In addition, SB202190 did not inhibit PKA activity.

X. Example V: Kinase Assays on a Glass Surface

1. Glass slides (Fisher, USA) were soaked in 28-30% ammonium hydroxide overnight at room temperature ("RT") with shaking.

2. The slides were rinsed with ultra-pure water four times for 5 minutes ("min") each, then rinsed with a large volume of 100% ethanol ("EtOH") to completely remove the water. Slides were then rinsed with 95% ethanol three times.

3. The slides were immersed in 1% 3-glycidoxypropyltrimethoxysilane (GPST) solution in 95% EtOH, 16 mM acetic acid ("HOAc") with shaking for 1 hr at room temperature. The slides were rinsed with 95% ethanol three times at RT.

4. The slides were cured at 135°C for 2 hrs under vacuum. After cooling, the slides can be stored in Argon for months before use.

5. Approximately 10 μl of each protein substrate (in 40% glycerol) were arrayed onto a 96-well PCR plate on ice. A manual spotting device (V&P Scientific, USA) was used to spot approximately 3 nl of each of the samples onto the GPTS-treated glass slide at RT. In one embodiment, 768 samples are spotted on a single slide. The slides were incubated in a covered and clean chamber at RT for one hour.

6. A slide was blocked with 10 ml blocking buffer (100 mM glycine, 100 mM Tris, pH 8.0, 50 mM NaCl) at RT for one hour. The slides were washed with TBS buffer (50 mM Tris, pH 8.0, 150 mM NaCl) three times and spun to dryness at 1500 rpm for 5 min.

7.   The substrate surfaces on the slides were covered with the HybriWell
Sealing System (Schleicher & Schuell, Germany) and 40 µl of kinase mixture, containing a
protein kinase and ^{33}P-γ-ATP as a labeling reagent, was added to the substrates on ice.

8.   The reaction was incubated at 30°C for 30 min in a humidity chamber. The
5   seals were peeled from the slides, and the slides immersed into large volume of PBS buffer
containing 50 mM EDTA. The slides were further washed with the same buffer 3 x 15 min
at RT. The washed slides were then dried with Kimwipes.

9.   To acquire the signals, the slides were exposed to a Phosphoimager screen
and the data analyzed using ImageQuant software.
10

XI.   References Cited

All references cited herein are incorporated herein by reference in their entirety and
15 for all purposes to the same extent as if each individual publication or patent or patent
application was specifically and individually indicated to be incorporated by reference in its
entirety for all purposes.

Many modifications and variations of this invention can be made without departing
from its spirit and scope, as will be apparent to those skilled in the art. The specific
20 embodiments described herein are offered by way of example only, and the invention is to
be limited only by the terms of the appended claims, along with the full scope of
equivalents to which such claims are entitled.

25

30

35

We claim:

1. A positionally addressable array comprising a plurality of different substances, selected from the group consisting of proteins, molecules comprising functional domains of said proteins, whole cells, and protein-containing cellular material, on a solid support, with each different substance being at a different position on the solid support, wherein the plurality of different substances consists of at least 100 different substances per cm^2.

2. The array of claim 1 wherein the plurality of different substances consists of between 100 and 1,000 different substances per cm^2.

3. The array of claim 1 wherein the plurality of different substances consists of between 1,000 and 10,000 different substances per cm^2.

4. The array of claim 1 wherein the plurality of different substances consists of between 10,000 and 100,000 different substances per cm^2.

5. The array of claim 1 wherein the plurality of different substances consists of between 100,000 and 1,000,000 different substances per cm^2.

6. The array of claim 1 wherein the plurality of different substances consists of between 1,000,000 and 10,000,000 different substances per cm^2.

7. The array of claim 1 wherein the plurality of different substances consists of between 10,000,000 and 25,000,000 different substances per cm^2.

8. The array of claim 1 wherein the plurality of different substances consists of at least 25,000,000 different substances per cm^2.

9. The array of claim 1 wherein the plurality of different substances consists of at least 10,000,000,000 different substances per cm^2.

10. The array of claim 1 wherein the plurality of different substances consists of at least 10,000,000,000,000 different substances per cm^2.

11. The array of claim 1 wherein the solid support is a glass slide.

12. The array of claim 1 wherein each different substance is present in a different well on the surface of the solid support.

13. The array of claim 12 wherein each different substance in a different well is bound to the surface of the solid support.

14. The array of claim 12 wherein each different substance in a different well is not bound to the surface of the solid support.

15. The array of claim 12 wherein each different substance in a different well is in solution.

- 47 -

16. The array of claim 12 wherein each well contains reagents for assaying biological activity of a protein or molecule.

17. A positionally addressable array comprising a plurality of different proteins, or molecules comprising functional domains of said proteins, on a solid support, with each different protein or molecule being at a different position on the solid support, wherein the plurality of proteins or molecules consists of at least 50% of all expressed proteins with the same type of biological activity in the genome of an organism.

18. The array of claim 17 wherein the plurality of proteins or molecules consists of at least 75% of all expressed proteins with the same type of biological activity in the genome of an organism.

19. The array of claim 17 wherein the plurality of proteins or molecules consists of at least 90% of all expressed proteins with the same type of biological activity in the genome of an organism.

20. The array of claim 17 wherein the organism is selected from the group consisting of bacteria, yeast, insects, and mammals.

21. The array of claim 17 wherein the expressed proteins with a biological activity of interest are selected from the group consisting of kinases, phosphatases, proteases, glycosidases, acetylases, other group transferring enzymes, nucleic acid binding proteins, hormone binding proteins, and DNA binding proteins.

22. A positionally addressable array comprising a plurality of different substances selected from the group consisting of proteins, molecules comprising functional domains of said proteins, whole cells, and protein-containing cellular material, on a solid support, with each different substance being at a different position on the solid support, wherein the solid support is selected from the group consisting of ceramics, amorphous silicon carbide, castable oxides, polyimides, polymethylmethacrylates, polystyrenes and silicone elastomers.

23. The array of claim 22 wherein the solid support is silicone elastomer.

24. The array of claim 23 wherein the solid support is polydimethylsiloxane.

25. A positionally addressable array comprising a plurality of different substances, selected from the group consisting of proteins, molecules comprising functional domains of said proteins, whole cells, and protein-containing cellular material, on a solid support, with each different substance being at a different position on the solid support, wherein the plurality of different substances are attached to the solid support via a 3-glycidooxypropyltrimethoxysilane linker.

26. An array comprising a plurality of wells on the surface of a solid support wherein the density of the wells is at least 100 wells/cm².

27. The array of claim 26 wherein the density of the wells is between 100 and 1,000 wells/cm².

5 28. The array of claim 26 wherein the density of the wells is between 1,000 and 10,000 wells/cm².

29. The array of claim 26 wherein the density of the wells is between 10,000 and 100,000 wells/cm².

30. The array of claim 26 wherein the density of the wells is between 100,000 and
10 1,000,000 wells/cm².

31. The array of claim 26 wherein the density of the wells is between 1,000,000 and 10,000,000 wells/cm².

32. The array of claim 26 wherein the density of the wells is between 10,000,000 and 25,000,000 wells/cm².

15 33. The array of claim 26 wherein the density of the wells is at least 25,000,000 wells/cm².

34. The array of claim 26 wherein the density of the wells is at least 10,000,000,000 wells/cm².

35. The array of claim 26 wherein the density of the wells is at least 10,000,000,000,000
20 wells/cm².

36. The array of claim 26 wherein a plurality of different substances, selected from the group consisting of proteins, molecules comprising functional domains of said proteins, whole cells, and protein-containing cellular material, is present in the wells, with each different substance being present in a different well.

25 37. The array of claim 36 wherein each different substance in a different well is bound to the surface of the solid support.

38. The array of claim 37 wherein each different substance in a different well is covalently bound to the surface of the solid support.

39. The array of claim 38 wherein each different substance in a different well is
30 covalently bound to the surface of the solid support through a linker.

40. The array of claim 39 wherein the linker is 3-glycidooxypropyltrimethoxysilane.

41. The array of claim 36 wherein each different substance in a different well is non-covalently bound to the surface of the solid support.

42. The array of claim 36 wherein each different substance in a different well is free of
35 binding to the surface of the solid support.

43. The array of claim 36 wherein each different substance in a different well is in solution.

44. The array of claim 26 wherein each well contains reagents for assaying biological activity of a protein or molecule.

5   45. The array of claim 26 wherein volumes of the wells are between 1 pl and 5 μl.

46. The array of claim 26 wherein volumes of the wells are between 1 nl and 1 μl.

47. The array of claim 26 wherein volumes of the wells are between 100 nl and 300 nl.

48. The array of claim 26 wherein the bottoms of the wells are square, round, V-shaped or U-shaped.

10   49. A method of making a positionally addressable array comprising a plurality of wells on the surface of a solid support comprising the step of:

    casting an array from a microfabricated mold designed to produce a density of wells on a solid surface of greater than 100 wells/cm².

50. A method of making a positionally addressable array comprising a plurality of wells

15   on the surface of a solid support comprising the steps of:

    (a)   casting a secondary mold from a microfabricated mold designed to produce a density of wells on a solid surface of greater than 100 wells/cm²; and

    (b)   casting at least one array from the secondary mold.

51. The method of claims 49 or 50 wherein the casting of an array further comprises the

20   steps of:

    (a)   covering the mold with a liquid cast material; and

    (b)   curing the cast material until the cast is solid.

52. The method of any of claims 49-51 wherein the density of the wells is between 100 and 1,000 wells/cm².

25   53. The method of any of claims 49-51 wherein the density of the wells is between 1,000 and 10,000 wells/cm².

54. The method of any of claims 49-51 wherein the density of the wells is between 10,000 and 100,000 wells/cm².

55. The method of any of claims 49-51 wherein the density of the wells is between

30   100,000 and 1,000,000 wells/cm².

56. The method of any of claims 49-51 wherein the density of the wells is between 1,000,000 and 10,000,000 wells/cm².

57. The method of any of claims 49-51 wherein the density of the wells is between 10,000,000 and 25,000,000 wells/cm².

35   58. The method of any of claims 49-51 wherein the density of the wells is greater than 25,000,000 wells/cm².

59. The method of any of claims 49-51 wherein the density of the wells is greater than 10,000,000,000 wells/cm^2.

60. The method of any of claims 49-51 wherein the density of the wells is greater than 10,000,000,000,000 wells/cm^2.

61. The method of claim 49 or 50 wherein the array is cast from silicone elastomer.

62. The method of claim 49 or 50 wherein the array is cast from polydimethylsiloxane.

63. The method of claim 51 wherein the liquid cast material is a silicone elastomer.

64. The method of claim 51 wherein the liquid cast material is polydimethylsiloxane.

65. A method of using a positionally addressable array comprising a plurality of different substances, selected from the group consisting of proteins, molecules comprising functional domains of said proteins, whole cells, and protein-containing cellular material, on a solid support, with each different substance being at a different position on the solid support, wherein the plurality of different substances consists of at least 100 different substances per cm^2, comprising the steps of:

    (a)   contacting a probe with the array; and

    (b)   detecting protein/probe interaction.

66. A method of using a positionally addressable array comprising a plurality of different proteins or molecules comprising functional domains of said proteins, on a solid support, with each different protein or molecule being at a different position on the solid support, wherein the plurality of proteins or molecules consists of at least 50% of all expressed proteins with the same type of biological activity in the genome of an organism, comprising the steps of:

    (a)   contacting a probe with the array; and

    (b)   detecting protein/probe interaction.

67. A method of using a positionally addressable array comprising a plurality of different substances, selected from the group consisting of proteins, molecules comprising functional domains of said proteins, whole cells, and protein-containing cellular material, on a solid support, with each different substance being at a different position on the solid support, wherein the solid support is selected from the group consisting of ceramics, amorphous silicon carbide, castable oxides, polyimides, polymethylmethacrylates, polystyrenes and silicone elastomers, comprising the steps of:

    (a)   contacting a probe with the array; and

    (b)   detecting protein/probe interaction.

68. A method of using a positionally addressable array comprising a plurality of different substances, selected from the group consisting of proteins, molecules

comprising functional domains of said proteins, whole cells, and protein-containing cellular material, on a solid support, with each different substance being at a different position on the solid support, wherein the plurality of different substances are attached to the solid support via a 3-glycidooxypropyltrimethoxysilane linker, comprising the steps of:

    (a)   contacting a probe with the array; and

    (b)   detecting protein/probe interaction.

69.   The method of any of claims 65-68 wherein the probe is an enzyme substrate or inhibitor.

70.   The method of claim 69 wherein the probe is a substrate or inhibitor of an enzyme chosen from the group consisting of kinases, phosphatases, proteases, glycosidases, acetylases, and other group transferring enzymes.

71.   The method of any of claims 65-68 wherein the probe is chosen from the group consisting of proteins, oligonucleotides, polynucleotides, DNA, RNA, small molecule substrates, drug candidates, receptors, antigens, steroids, phospholipids, antibodies, glutathione, immunoglobulin domains, maltose, nickel, dihydrotrypsin, and biotin.

72.   The method of any of claims 65-68 wherein detection of protein/probe interaction is via mass spectrometry.

73.   The method of any of claims 65-68 wherein detection of protein/probe interaction is via a method chosen from the group consisting of chemiluminescence, fluorescence, radiolabeling, and atomic force microscopy.

74.   A method of using a positionally addressable array comprising the steps of:

    (a)   depositing a plurality of different substances, selected from the group consisting of proteins, molecules comprising functional domains of said proteins, whole cells, and protein-containing cellular material, on a solid support, with each different substance being at a different position on the solid support, wherein the plurality of different substances consists of at least 100 different substances per cm^2;

    (b)   contacting a probe with the array; and

    (c)   detecting protein/probe interaction.

75.   The method of claim 74 wherein the solid support is a glass slide.

76.   A method of using a positionally addressable array comprising the steps of:

    (a)   depositing a plurality of different proteins, or molecules comprising functional domains of said proteins, on a solid support, with each different protein or molecule being at a different position on the solid

support, wherein the plurality of proteins or molecules consists of at least 50% of all expressed proteins with the same type of biological activity in the genome of an organism;

(b)   contacting a probe with the array; and

5   (c)   detecting protein/probe interaction.

77.   The method of claim 76 wherein the solid support is a glass slide.

78.   A method of making a positionally addressable array comprising the steps of:

(a)   casting an array from a microfabricated mold designed to produce a density of wells on a solid surface of greater than 100 wells/cm^2; and

10   (b)   depositing in wells a plurality of different substances, selected from the group consisting of proteins, molecules comprising functional domains of said proteins, whole cells, and protein-containing cellular material, on a solid support, with each different substance being in a different well on the solid support.

15 79.   A method of making a positionally addressable array comprising the steps of:

(a)   casting a secondary mold from a microfabricated mold designed to produce a density of wells on a solid surface of greater than 100 wells/cm^2;

(b)   casting at least one array from the secondary mold; and

20   (c)   depositing in wells a plurality of different substances, selected from the group consisting of proteins, molecules comprising functional domains of said proteins, whole cells, and protein-containing cellular material, on a solid support, with each different substance being in a different well on the solid support.

25 80.   A method of identifying an antigen that activates a lymphocyte comprising the steps of:

(a)   contacting a positionally addressable array with a plurality of lymphocytes, said array comprising a plurality of potential antigens on a solid support, with each different antigen being at a different position on the solid support, wherein the density of different antigens is at least 100 different antigens per cm^2; and

(b)   detecting positions on the solid support where lymphocyte activation occurs.

81.   The method of Claim 80 wherein the lymphocytes are derived from a patient.

35

- 53 -

82. The method of Claim 80 wherein the antigens are selected from the group consisting of antigens of pathogens, antigens of autologous tissues, tissue-specific antigens, disease-specific antigens, and synthetic antigens.

83. The method of Claim 80 wherein lymphocyte activation is detected by measuring antibody synthesis.

84. The method of Claim 80 wherein lymphocyte activation is detected by measuring the incorporation of ^{3}H-thymidine by a lymphocyte.

85. The method of Claim 80 wherein lymphocyte activation is detected by determining the expression of cell surface molecules induced or suppressed by lymphocyte activation.

86. The method of Claim 80 wherein lymphocyte activation is detected by determining the expression of secreted molecules induced by lymphocyte activation.

87. The method of Claim 80 wherein lymphocyte activation is detected by measuring the release of ^{51}chromium.

88. A method of determining the specificity of an antibody preparation comprising the steps of:

    (a)   contacting a positionally addressable array with an antibody preparation, said array comprising a plurality of potential antigens on a solid support, with each different antigen being at a different position on the solid support, wherein the density of different antigens is at least 100 different antigens per cm^2; and

    (b)   detecting positions on the solid support where binding by an antibody in said antibody preparation occurs.

89. The method of Claim 88 wherein the antibody preparation comprises antiserum, a monoclonal antibody, or a polyclonal antibody.

90. The method of Claim 88 wherein the antibody preparation comprises Fab fragments, chimeric, single chain, humanized, or synthetic antibodies.

91. The method of Claim 88 wherein antibody binding is detected by contacting the array with a fluorescently labeled secondary antibody that binds to antibody in said antibody preparation; removing unbound secondary antibody; and detecting bound label on the array.

92. A method of identifying a mitogen comprising the steps of:

    (a)   contacting a positionally addressable array with a population of cells; said array comprising a plurality of different substances, selected from the group consisting of proteins, molecules comprising functional domains of said proteins, whole cells, and protein-

- 54 -

NY2 - 1102584.9

containing cellular material, on a solid support, with each different substance being at a different position on the solid support, wherein the density of different substances is at least 100 different substances per cm²; and

5   (b)   detecting positions on the solid support where mitogenic activity is induced in a cell.

93.   A kit comprising:

      (a)   one or more arrays comprising a plurality of wells on the surface of a solid support wherein the density of the wells is at least 100 wells/cm²; and

10

      (b)   in one or more containers, one or more probes, reagents, or other molecules.

94.   The kit according to Claim 93 wherein said one or more containers comprise a reagent useful for assaying biological activity of a protein.

15 95.   The kit according to Claim 93 wherein said one or more containers comprise a reagent useful for assaying interactions between a probe and a protein.

96.   The kit according to Claim 94 or 95 wherein the reagent is in solution.

97.   The kit according to Claim 94 or 95 wherein the reagent is in solid form.

98.   The kit according to Claim 94 or 95 wherein the reagent is contained in each well of

20   the array.

99.   The kit according to Claim 94 or 95 wherein the reagent is contained in selected wells of the array.

100.   The kit according to Claim 93 wherein said one or more containers contain a solution reaction mixture for assaying biological activity of a protein or molecule.

25 101.   The kit according to Claim 100 wherein said one or more containers contain one or more substrates to assay said biological activity.

102.   A kit comprising:

      (a)   one or more positionally addressable arrays comprising a plurality of different substances, selected from the group consisting of proteins,

30   molecules comprising functional domains of said proteins, whole cells, and protein-containing cellular material, on a solid support, with each different substance being at a different position on the solid support, wherein the plurality of different substances consists of at least 100 different substances per cm²; and

35   (b)   in one or more containers, one or more probes, reagents, or other molecules.

103. The kit according to Claim 102 wherein the substances are attached to the surface of wells on the solid support.

104. The kit according to Claim 103 wherein the substances are proteins, and the proteins are at least 50% of all expressed proteins with the same type of biological activity in an organism.

105. The kit according to Claim 104 wherein the substances are proteins or molecules comprising functional domains of said proteins, and the proteins or molecules are selected from the group consisting of kinases, phosphatases, proteases, glycosidases, acetylases, nucleic acid binding proteins, and hormone binding proteins.

ABSTRACT

The present invention relates to protein chips useful for the large-scale study of protein function where the chip contains densely packed reaction wells. The invention also

5   relates to methods of using protein chips to assay simultaneously the presence, amount, and/or function of proteins present in a protein sample or on one protein chip, or to assay the presence, relative specificity, and binding affinity of each probe in a mixture of probes for each of the proteins on the chip. The invention also relates to methods of using the protein chips for high density and small volume chemical reactions. Also, the invention

10   relates to polymers useful as protein chip substrates and methods of making protein chips. The invention further relates to compounds useful for the derivatization of protein chip substrates.

15

20

25

30

35

Exhibit 3

- 3 -   Snyder *et al.*
                                                09/849,781

Amendments to the Claims

This listing of claims will replace all prior versions, and listings of claims in the application.

Claim 1.   (Currently amended) A positionally addressable array comprising a plurality of different substances on a solid support, with each different substance being at a different position on the solid support, wherein the density of the different substances on the solid support is at least 100 different substances per cm^2, and wherein the plurality of different substances comprises at least 61 purified active kinases or functional kinase domains thereof of a mammal, 61 purified active kinases or functional kinase domains thereof of a yeast, or 61 purified active kinases or functional kinase domains thereof of a Drosophila.

Claim 2.   (Previously presented) The array of claim 1 wherein the density of the different substances on the array is between 100 and 1,000 different substances per cm^2.

Claim 3.   (Previously presented) The array of claim 1 wherein the density of the different substances on the array is between 1,000 and 10,000 different substances per cm^2.

Claim 4.   (Previously presented) The array of claim 1 wherein the density of the different substances on the array is between 10,000 and 100,000 different substances per cm^2.

Claim 5.   (Previously presented) The array of claim 1 wherein the density of the different substances on the array is between 100,000 and 1,000,000 different substances per cm^2.

Claim 6.   (Previously presented) The array of claim 1 wherein the density of the different substances on the array is between 1,000,000 and 10,000,000 different substances per cm^2.

Claim 7.   (Previously presented) The array of claim 1 wherein the density of the different substances on the array is between 10,000,000 and 25,000,000 different substances per cm^2.

Claim 8.   (Previously presented) The array of claim 1 wherein the density of the different substances on the array is at least 25,000,000 different substances per cm^2.

Claim 9.   (Previously presented) The array of claim 1 wherein the density of the different substances on the array is at least 10,000,000,000 different substances per cm^2.

Claim 10.   (Previously presented) The array of claim 1 wherein the density of the different substances on the array is at least 10,000,000,000,000 different substances per cm^2.

Claim 11.   (Original) The array of claim 1 wherein the solid support is a glass slide.

Claim 12.   (Withdrawn) The array of claim 1 wherein each different substance is present in a different well on the surface of the solid support.

Claim 13.   (Withdrawn) The array of claim 12 wherein each different substance in a different well is bound to the surface of the solid support.

Claim 14.   (Withdrawn) The array of claim 12 wherein each different substance in a different well is not bound to the surface of the solid support.

Claim 15.   (Withdrawn) The array of claim 12 wherein each different substance in a different well is in solution.

Claim 16.   (Withdrawn) The array of claim 12 wherein each well contains reagents for assaying biological activity of a protein or molecule.

Claims 17-92. (Canceled).

Claim 93.   (Withdrawn) A kit comprising:

(a)   one or more arrays of claim 1 comprising a plurality of wells on the surface of the solid support wherein the density of the wells is at least 100 wells/cm^2, wherein each of said different substances is present in a different well; and

(b)   in one or more containers, one or more probes, reagents, or other second molecules.

Claim 94.   (Withdrawn) The kit according to claim 93 wherein said one or more containers comprise a reagent useful for assaying biological activity of a protein.

Claim 95.   (Withdrawn) The kit according to claim 93 wherein said one or more containers comprise a reagent useful for assaying interactions between a probe and a protein.

Claim 96.   (Withdrawn) The kit according to claim 94 or 95 wherein the reagent is in solution.

Claim 97.   (Withdrawn) The kit according to claim 94 or 95 wherein the reagent is in solid form.

Claim 98.   (Withdrawn) The kit according to claim 94 or 95 wherein the reagent is contained in each well of the array.

Claim 99.   (Withdrawn) The kit according to claim 94 or 95 wherein the reagent is contained in selected wells of the array.

Claim 100.   (Withdrawn) The kit according to claim 93 wherein said one or more containers contain a solution reaction mixture for assaying biological activity.

Claim 101.   (Withdrawn) The kit according to claim 100 wherein said one or more containers contain one or more substrates to assay said biological activity.

Claims 102-105.   (Canceled).

Claim 106.   (Withdrawn) The array of claim 1 wherein the solid support is composed of a silicone elastomeric material.

Claim 107.   (Withdrawn) The array of claim 106 wherein the silicone elastomeric material is polydimethylsiloxane.

Claims 108 to 111.   (Canceled).

Claim 112.   (Withdrawn) The kit of claim 93 wherein the solid support is selected from the group consisting of a ceramic, amorphous silicon carbide, castable oxide, polyimide, polymethylmethacrylate, polystyrene, and silicone elastomer.

Claim 113.   (Withdrawn) The kit of claim 112 wherein the solid support is a silicone elastomer.

Claim 114.   (Withdrawn) The kit of claim 112 wherein the solid support is a polydimethylsiloxane.

Claim 115.   (Withdrawn) The kit of claim 93 wherein the plurality of different substances are attached to the solid support via a 3-glycidoxypropyltrimethoxysilane linker.

Claim 116.   (Withdrawn) The kit of claim 93 wherein the density of the wells is between 100 and 1,000 wells/cm^2.

Claim 117.   (Withdrawn) The kit of claim 93 wherein the density of the wells is between 1,000 and 10,000 wells/cm^2.

Claim 118.   (Withdrawn) The kit of claim 93 wherein the density of the wells is between 10,000 and 100,000 wells/cm^2.

Claim 119.   (Withdrawn) The kit of claim 93 wherein the density of the wells is between 100,000 and 1,000,000 wells/cm^2.

Claim 120.   (Withdrawn) The kit of claim 93 wherein the density of the wells is between 1,000,000 and 10,000,000 wells/cm^2.

Claim 121.   (Withdrawn) The kit of claim 93 wherein the density of the wells is between 10,000,000 and 25,000,000 wells/cm^2.

Claim 122.   (Withdrawn) The kit of claim 93 wherein each different substance in a different well is bound to the surface of the solid support.

Claim 123.   (Withdrawn) The kit of claim 122 wherein each different substance in a different well is covalently bound to the surface of the solid support.

Claim 124.   (Withdrawn) The kit of claim 123 wherein each different substance in a different well is covalently bound to the surface of the solid support through a linker.

Claim 125.   (Withdrawn) The kit of claim 124 wherein the linker is 3-glycidoxypropyltrimethoxysilane.

Claim 126.   (Withdrawn) The kit of claim 122 wherein each different substance in a different well is non-covalently bound to the surface of the solid support.

Claim 127.   (Withdrawn) The kit of claim 93 wherein each different substance in a different well is free of binding to the surface of the solid support.

Claim 128.   (Withdrawn) The kit of claim 93 wherein each different substance in a different well is in solution.

Claim 129.   (Withdrawn) The kit of claim 93 wherein each well contains reagents for assaying biological activity.

Claim 130.   (Withdrawn) The kit of claim 93 wherein volumes of the wells are between 1 pl and 5μl.

Claim 131.   (Withdrawn) The kit of claim 93 wherein volumes of the wells are between 1 nl and 1 μl.

Claim 132.   (Withdrawn) The kit of claim 93 wherein volumes of the wells are between 100 nl and 300 nl.

Claim 133.   (Withdrawn) The kit of claim 93 wherein the bottoms of the wells are square, round, V-shaped or U-shaped.

Claims 134-137.   (Canceled).

Claim 138.   (Withdrawn) The array of claim 1 wherein the solid support is selected from the group consisting of a ceramic, amorphous silicon carbide, castable oxide, polyimide, polymethylmethacrylate, polystyrene, and silicone elastomer.

Claim 139. (Withdrawn) The array of claim 1 wherein the solid support is a silicone elastomer.

Claim 140. (Withdrawn) The array of claim 139 wherein the solid support is a polydimethylsiloxane.

Claim 141. (Previously presented) The array of claim 1 wherein the plurality of different substances are attached to the solid support via a 3-glycidoxypropyl-trimethoxysilane linker.

Claim 142. (Withdrawn) The array of claim 12 wherein the density of the wells is between 100 and 1,000 wells/cm^2.

Claim 143. (Withdrawn) The array of claim 12 wherein the density of the wells is between 1,000 and 10,000 wells/cm^2.

Claim 144. (Withdrawn) The array of claim 12 wherein the density of the wells is between 10,000 and 100,000 wells/cm^2.

Claim 145. (Withdrawn) The array of claim 12 wherein the density of the wells is between 100,000 and 1,000,000 wells/cm^2.

Claim 146. (Withdrawn) The array of claim 12 wherein the density of the wells is between 1,000,000 and 10,000,000 wells/cm^2.

Claim 147. (Withdrawn) The array of claim 12 wherein the density of the wells is between 10,000,000 and 25,000,000 wells/cm^2.

Claim 148. (Withdrawn) The array of claim 12 wherein each different substance in a different well is bound to the surface of the solid support.

Claim 149. (Withdrawn) The array of claim 148 wherein each different substance in a different well is covalently bound to the surface of the solid support.

Claim 150.   (Withdrawn) The array of claim 149 wherein each different substance in a different well is covalently bound to the surface of the solid support through a linker.

Claim 151.   (Withdrawn) The array of claim 150 wherein the linker is 3-glycidoxypropyltrimethoxysilane.

Claim 152.   (Withdrawn) The array of claim 148 wherein each different substance in a different well is non-covalently bound to the surface of the solid support.

Claim 153.   (Withdrawn) The array of claim 12 wherein each different substance in a different well is free of binding to the surface of the solid support.

Claim 154.   (Withdrawn) The array of claim 12 wherein each different substance in a different well is in solution.

Claim 155.   (Withdrawn) The array of claim 12 wherein each well contains reagents for assaying biological activity.

Claim 156.   (Withdrawn) The array of claim 12 wherein volumes of the wells are between 1 pl and 5 µl.

Claim 157.   (Withdrawn) The array of claim 12 wherein volumes of the wells are between 1 nl and 1 µl.

Claim 158.   (Withdrawn) The array of claim 12 wherein volumes of the wells are between 100 nl and 300 nl.

Claim 159.   (Withdrawn) The array of claim 12 wherein the bottoms of the wells are square, round, V-shaped or U-shaped.

Claims 160-161.   (Canceled).

Claim 162.   (Withdrawn) The kit of claim 93 wherein the organism is selected from the group consisting of human, primate, mouse, rat, cat, dog, horse, and cow.

Claim 163.   (Canceled).

Claim 164.   (Currently amended) The array of claim 1 wherein the ~~mammal~~ organism is selected from the group consisting of human, primate, mouse, rat, cat, dog, horse, and cow.

Claim 165.   (Withdrawn) The array of claim 12 wherein the organism is selected from the group consisting of human, primate, mouse, rat, cat, dog, horse, and cow.

Claim 166.   (Canceled).

Claim 167.   (Withdrawn) The kit of claim 162, wherein the organism is human.

Claim 168.   (Canceled).

Claim 169.   (Previously presented) The array of claim 164, wherein the organism is human.

Claim 170.   (Canceled).

Claim 171.   (Withdrawn) The kit of claim 162, wherein the organism is mouse.

Claim 172.   (Canceled).

Claim 173.   (Previously presented) The array of claim 164, wherein the organism is mouse.

Claim 174.   (Currently amended) The array of claim 164 [[166]], wherein the organism is mouse.

Claim 175.   (Withdrawn) The kit of claim 162, wherein the organism is rat.

Claim 176.   (Canceled).

Claim 177.   (Previously presented) The array of claim 164, wherein the organism is rat.

Claims 178-180.   (Canceled).

Claim 181.   (Previously presented) The positionally addressable protein array of claim 1, wherein the plurality of different substances comprises 61 different purified active kinases of an organism.

Claim 182.   (Currently amended) The positionally addressable protein array of claim 1, wherein the plurality of different substances comprises 92 different purified active kinases of a mammal, a yeast, or a Drosophila an organism.

Claim 183.   (Currently amended) The positionally addressable protein array of claim 1, wherein the plurality of different substances comprises 110 different purified active kinases of a mammal, a yeast, or a Drosophila an organism.

Claim 184.   (Currently amended) The positionally addressable protein array of claim 1, wherein the plurality of different substances comprises 116 different purified active kinases of a mammal, a yeast, or a Drosophila an organism.

Claim 185.   (Currently amended) The positionally addressable protein array of claim 1, wherein the plurality of different substances comprises 119 different purified active kinases of a mammal, a yeast, or a Drosophila an organism.

Claim 186.   (Currently amended) The positionally addressable protein array of claim 1, wherein the plurality of different substances comprises 122 purified active different kinases of a mammal, a yeast, or a Drosophila an organism.

Claim 187.   (Canceled).

Claim 188.   (Previously presented) The positionally addressable array of claim 1, wherein the kinases are yeast kinases.

Claims 189-192.   (Canceled).

Claim 193.   (Previously presented) The positionally addressable array of claim 1, wherein the different substances are 61 purified active kinases.

Claim 194.   (Currently amended) The positionally addressable array of claim 193, wherein the kinases are ~~members of the~~ serine/threonine kinase family members, ~~members of the~~ tyrosine kinase family members, or ~~the kinases are members of the~~ serine/threonine kinase ~~family~~ and ~~members of the~~ tyrosine kinase family members.

Claim 195.   (Currently amended) The positionally addressable array of claim 1, wherein the functional kinase domains are functional kinase domains of ~~members of~~ ~~the~~ serine/threonine kinase family members, functional kinase domains of ~~members of~~ ~~the~~ tyrosine kinase family members, or ~~wherein the~~ functional kinase domains ~~comprise~~ ~~functional kinase domains~~ of ~~kinases that are members of the~~ serine/threonine kinase family members and functional kinase domains of ~~kinases that are members of the~~ tyrosine kinase family members.

Claim 196.   (Withdrawn) The positionally addressable array of claim 1, wherein the kinases or functional kinase domains are recombinant proteins.

Claim 197.   (Withdrawn) The positionally addressable array of claim 196, wherein the recombinant proteins are recombinant fusion proteins.

Claim 198.   (Canceled).

Exhibit 4

UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
09/849,781	05/04/2001	Michael Snyder	2493.0010002/RWE/JKM	9891

26111   7590   07/07/2009
STERNE, KESSLER, GOLDSTEIN & FOX P.L.L.C.
1100 NEW YORK AVENUE, N.W.
WASHINGTON, DC 20005

EXAMINER
WESSENDORF, TERESA D.

ART UNIT	PAPER NUMBER
1639	

MAIL DATE	DELIVERY MODE
07/07/2009	PAPER

Please find below and/or attached an Office communication concerning this application or proceeding.

The time period for reply, if any, is set in the attached communication.

PTOL-90A (Rev. 04/07)

	Application No.	Applicant(s)
Office Action Summary	09/849,781	SNYDER ET AL.
	Examiner	Art Unit	
	TERESA WESSENDORF	1639	

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --

Period for Reply

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE _3_ MONTH(S) OR THIRTY (30) DAYS, WHICHEVER IS LONGER, FROM THE MAILING DATE OF THIS COMMUNICATION.
- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133).
- Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

Status

1)☐ Responsive to communication(s) filed on _____.

2a)☐ This action is **FINAL**.   2b)☒ This action is non-final.

3)☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

Disposition of Claims

4)☒ Claim(s) _See Continuation Sheet_ is/are pending in the application.

    4a) Of the above claim(s) _See Continuation Sheet_ is/are withdrawn from consideration.

5)☐ Claim(s) _____ is/are allowed.

6)☒ Claim(s) _1-11, 141, 164, 169, 173, 174, 177, 181-186, 188 and 193-195_ is/are rejected.

7)☐ Claim(s) _____ is/are objected to.

8)☐ Claim(s) _____ are subject to restriction and/or election requirement.

Application Papers

9)☐ The specification is objected to by the Examiner.

10)☐ The drawing(s) filed on _____ is/are: a)☐ accepted or b)☐ objected to by the Examiner.

    Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).

    Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).

11)☐ The oath or declaration is objected to by the Examiner. Note the attached Office Action or form PTO-152.

Priority under 35 U.S.C. § 119

12)☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).

    a)☐ All b)☐ Some * c)☐ None of:

      1.☐ Certified copies of the priority documents have been received.

      2.☐ Certified copies of the priority documents have been received in Application No. _____.

      3.☐ Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).

    * See the attached detailed Office action for a list of the certified copies not received.

Attachment(s)

1)☐ Notice of References Cited (PTO-892)

2)☐ Notice of Draftsperson's Patent Drawing Review (PTO-948)

3)☒ Information Disclosure Statement(s) (PTO/SB/08)
    Paper No(s)/Mail Date _4/20/09_.

4)☐ Interview Summary (PTO-413)
    Paper No(s)/Mail Date. _____ .

5)☐ Notice of Informal Patent Application

6)☐ Other: _____.

Continuation of Disposition of Claims: Claims pending in the application are 1-16,93-101,106,107,112-133,138-159,162,164,165,167,169,171,173-175,177,181-186,188 and 193-197.
Continuation of Disposition of Claims: Claims withdrawn from consideration are 12-16,93-101,106,107,112-133,138-140,142-159,162,165,167,171,175,196 and 197.

DETAILED ACTION

Continued Examination Under 37 CFR 1.114

A request for continued examination under 37 CFR 1.114, including the fee set forth in 37 CFR 1.17(e), was filed in this application after final rejection. Since this application is eligible for continued examination under 37 CFR 1.114, and the fee set forth in 37 CFR 1.17(e) has been timely paid, the finality of the previous Office action has been withdrawn pursuant to 37 CFR 1.114. Applicant's submission filed on 4/2/09 has been entered.

Claims Status

Claims 1-16, 93-101, 106, 107, 112-133, 138-159, 162, 164-165, 167, 169, 171, 173-175, 177, 181-186, 188, and 193-197 are pending.

Claims 17-92, 102-105, 108-111, 134-137, 160-161, 163, 166, 168, 170, 172, 176, 178-180, 187, 189-192 and 198 have been cancelled.

Claims 12-16, 93-101, 106, 107, 112-133, 138-140, 142-159, 162, 165, 167, 171, 175 and 196-197 are withdrawn from further consideration pursuant to 37 CFR 1.142(b), as being drawn to nonelected species, there being no allowable generic or linking claim.

Claims 1-11, 141, 164, 169, 173, 174, 177, 181-186, 188 and 193-195 are under consideration in this Office Action.

The inadvertent indication that claims 173-174, 177, 181-186, 188 and 192-195 are withdrawn from consideration at page 2 of the last Office action is regretted. These claims have been rejected throughout the Office action, as stated by applicants and the rejections are reiterated as shown below. Applicants' request that these claims be confirmed as not withdrawn from examination in the instant application is hereby granted.

Withdrawn Objection and Rejections

In view of applicants' arguments and amendments to the claims, the 35 USC 112, first paragraph (new matter); second paragraph and obviousness double patenting rejections have been withdrawn.

Claim Rejections - 35 USC § 112

Claims 1-11, 141, 164, 169, 173, 174, 177, 181-186, 188 and 193-195, as amended, are rejected under 35 U.S.C. 112, first paragraph, as failing to comply with the written description requirement. The claim(s) contains subject matter, which was not described in the specification in such a way as to reasonably convey to one skilled in the relevant art that the inventor(s), at the time the application was filed, had possession of the claimed invention.

A "written description of an invention involving a
chemical genus, like a description of a chemical species,
requires a precise definition, such as by structure, formula
[or] chemical name of the claimed subject matter sufficient to
distinguish it from other materials". University of California
v. Eli Lilly and Col, 43 USPQ 2d 1398, 1405(1997), quoting
Fiefs V. Revel, 25 USPQ 2d 1601m 16106 (Fed. Cir. 1993.

The claimed invention is drawn to a positionally
addressable array comprising a plurality of different
substances on a solid support, with each different substance
being at a different position on the solid support, wherein
the density of the different substances on the solid support
is at least 100 different substances per cm2, and wherein the
plurality of different substances comprises 61 purified active
kinases or functional kinase domains thereof of a mammal, 61
purified active kinases or functional kinase domains thereof
of a yeast, or 61 purified active kinases or functional kinase
domains thereof of a Drosophila.

The specification fails to provide an adequate written
description of 61 purified kinase and functional domains thereof
from any organism such as mammals, bacteria, viruses. The
specification provides general statements of these various

kinases in an organism which are not a detail description of the
invention. The detail description in the specification (Example
I, page 27) describes the 122 kinase genes specifically from the
yeast genome, not from the broad claimed any kinase from any
type of organisms or functional domains thereof. A written
description of a single species would not be a written
description for the genus as claimed. At the time of applicants'
invention kinases in any organism included in the huge scope of
the claim has not been fully characterized such that it has been
positioned in an array without denaturing the purified protein.
A skilled artisan recognizes that one cannot rule out the
possibility that kinases other than the desired enzyme can
contaminate any type of purification preparations.
Notwithstanding this, the kind/type and preparation of a
substrate compatible with the purified protein is also a factor
to consider for a purified protein to be active in an array.
Furthermore, "although most of the kinases were active in [our]
assays, several were not. Presumably, our preparations of these
latter kinases either lack sufficient quantities of an activator
or were not purified under activating conditions. For example,
Cdc28 which was not active in our assays, might be lacking its
activating cyclins. For the case of Hog1, cells were treated

with high salt to activate the enzyme…." (paragraph [0161] of

the instant specification, publication no. 20030207467).

Attention is also drawn to the numerous prior art cited by

applicants, inter alia, the Anderson reference, which teaches

the numerous unforeseeable factors of a purified kinase

positioned in an array.

Anderson states that:

> …protein microarrays have still not found widespread use,
> in part because producing them is challenging.
> Historically, it has required the high-throughput
> production and purification of protein, which then must be
> spotted on the arrays. Once printed, concerns remain about
> the shelf life of proteins on the arrays.

Shaw et al (Drug Discovery and Development, Exhibit B)

concords with the statement that:

> "[i]t was first thought that protein biochips would just be
> an extension of DNA microarrays, and that hasn't exactly
> panned out," says Bodovitz. That's because proteins have
> proven to be much trickier to work with in array format
> than their genomic counterparts. First of all, there are
> issues of stability. Membrane proteins, for example, make
> up the majority of potential drug targets, but they're
> particularly challenging to stabilize. Then there's the
> choice of immobilization technique, which determines how
> well the target protein presents itself to the capture
> agent, and the problem of nonspecific binding. And of
> course, proteins are inherently unstable outside their
> natural habitat of living cells, making them much more
> challenging than DNA to tag and manipulate.

An applicant shows possession of the claimed invention by

describing the claimed invention with all of its limitations

using such descriptive means as words, structures and formulas

to show that the invention is complete. Lockwood v. American

Airlines, Inc., 107 F.3d 1565, 1572, 41 USPQM 1961, 1966 (Fed.

Cir. 1997); MPEP 2163. Herein, kinase has been described only in

words. The characterization of the different kinases from one

organism to another from the numerous kinases and numerous

organisms has not been adequately described to distinguish one

from the other. To date only a few organisms are fully

characterized and the kinase region has not been fingerprinted

in a partly or even fully characterized gene. The description

lacks structural characterization of a purified kinase as

generically claim. It does not distinguish one kinase from

another and/or one organism from another positioned in any

kind/type of substrate array and reasonably expect the purified

kinase to retain its active form.

Claim Rejections - 35 USC § 112

Claims 1-11, 141, 164, 169, 173, 174, 177, 181-186, 188 and

193-195, as amended, are rejected under 35 U.S.C. 112, first

paragraph, because the specification, while being enabling for

yeast protein' kinases of the Ser/Thr and tyrosine kinase

family, does not reasonably provide enablement for the broad

scope of an array of 61 kinases and functional domain kinase

from an organism as mammal, yeast or Drosophila. The

specification does not enable any person skilled in the art to

which it pertains, or with which it is most nearly connected, to

make and use the invention commensurate in scope with these

claims for reasons as repeated below.

The claimed array comprises a broad genus of compositions.

The claimed different substances encompass any members of the

protein kinase from the organism of 'mammal, yeast, or

Drosophila' which is broader than the enabling disclosure. The

claimed array represents enormous scope because the claims do

not place any limitations on the kind, number and/or length of

kinase either singly from one family of organism or a

combination(s) from the different numerous recited organisms.

The instant specification is directed to an array comprising a

plurality of different yeast protein kinase, specifically 122

different yeast protein' kinases of the Ser/Thr and tyrosine

kinase family members (see specification: example I, pg. 27,

line 19 thru pg. 35, line 20; example II, pg. 41, line 19 thru

pg. 43, line 6). The specification does not provide reasonable

assurance to one skilled in the art that the 61 kinases found in

the yeast could be found in any or all of the organisms such as

mammals especially the functional domain thereof. It is not

apparent from the specification whether the same number of

kinases or the kind of kinases or functional domain thereof can
be found in any other organisms and made into an array. It is
not apparent from the disclosure as to the functional domain of
the kinase and the specific function attributed to said kinase
positioned on the array. The general knowledge and level of
skill in the art do not supplement the omitted description
because specific, not general, guidance is what is needed. In a
highly unpredictable art, as biotechnology, where one cannot
predict whether one species would be predictive to the huge
scope of the claim, one cannot make a priori statement without
any experimental studies. Factors such as the compatibility of
the array with the substrate and compounds disposed therein, the
compounds (kinases) itself and other unpredictable variables can
affect the active form of any kinase. Thus, one cannot predict
from a single species its correspondence or extrapolation to the
genus, as claimed.

Response to Arguments

Applicants note that the use of kinases from other
organisms, including mammals and Drosophila, in the arrays of
the presently claimed invention would not have required undue
experimentation, but rather, simple, straightforward
experiments. The protein kinases and functional kinase domains
for use in the presently claimed invention are all well-known,

well-characterized proteins that the ordinarily skilled artisan
would easily comprehend.

In reply, attention is drawn to the instant disclosure at
e.g., Example I which states that the tyrosine kinase family
members do not exist although seven protein kinases that
phophorylate have been reported. Applicants' arguments that
array from any organisms are simple and straightforward are
inconceivable given only the single species in the
specification.

Applicants rely on the Hanks reference for its disclosure
that "there are now hundreds of different members [of the kinase
superfamily] whose sequences are known." Hanks and Hunter, page
576. Furthermore, kinases, for example serine kinases, were
already readily recognized in 1995 by virtue of their conserved
subdomains. Applicants similarly rely on numerous references and
the Synder declaration to show that kinases are well-
characterized and known in the art.

In reply, there is nothing in Hanks' reference that
discloses these hundreds of kinases are from any mammals or
Drosophila or from any other origin as broadly claimed.

All of the references cited by applicants and the Synder
declaration provide also only general statements. The
characterization is for a specific kinase not in purified and

active form that has been positioned in an array for any kind of

organisms fully or partly characterized. Each of the references

and the declaration fail to take in consideration the numerous

factors of the claim genus array besides the characterization of

the kinase. For example, none of the references describes how

the numerous kinases from different mammals, different strains

of yeast or Drosophila can be purified. How this pure kinase has

been positioned in an array and still remains active.

Applicants' statements throughout the REMAKRS as to the

skepticism in the art provide evidence as to the high

unpredictability in the art. Furthermore, applicants in the

specification (Published patent 20030207467) states at e.g.,

paragraph [0038] FIG. 1b, that from three attempts, 106 kinase

proteins were purified. **In spite of repeated attempts,** the last

14 of 119 GST fusions were undetectable by immunoblotting

analysis. Further, at page 34 of the instant REMARKS, applicants

state:

> In Ge, "UPA, a universal protein array system for
> quantitative detection of protein-protein, protein-DNA,
> protein-RNA and protein-ligand interactions the author was
> only able to produce arrays comprising 48 proteins at a
> very low density, utilizing a traditional purification
> format. Extension of this disclosure to arrays comprising
> at least 100 different substances per cm2 would have
> required **extensive, undue experimentation** beyond the scope
> of the disclosure provided in this reference. (Emphasis
> added.)

Thus, an enabling disclosure for a single species of a

protein would not be enabling for the broad scope of other

protein kinases from any kind of organisms. [Reciting the kinase

is a Thr/Ser and Tyr of a yeast protein would overcome this

rejection.)

Claim Rejections - 35 USC § 112

Claims 1-11, 141, 164, 169, 173, 174, 177, 181-186, 188 and

193-195, as amended, are rejected under 35 U.S.C. 112, second

paragraph, as being indefinite for failing to particularly point

out and distinctly claim the subject matter which applicant

regards as the invention.

A. In claim 1, the metes and bounds of the claim

"functional kinase domains" is vague and indefinite as to the

kind, length or region the domain encompasses in a purified,

active form to be a functional kinase. It is not clear whether

the functional kinase domain is positioned together with the

full length kinase with the 61 different kinases from the

different organisms. And, still expect to be pure and active

without being masked by the full length kinases.

B. Non-sequitur for "the solid support" in claims 11 and

141. The base claim 1 does not recite a solid support. Also,

"the organism" in claim 164, claim 169, claim 173, 175 and 177:

"the serine/threonine kinase family", "the tyrosine kinase

family" in claims 194 and 195 all lack antecedent basis of

support from the base claim 1.

    C. Claim 174 depends on canceled claim 166.

    D. Claims 181-186 and 193 which each recite an organism

broaden the base claim 1. Claim 1 recites mammals, yeast and

Drosophila. However, organisms include e.g., bacteria, viruses

and other organisms besides the ones recited therein.

Furthermore, claim 1 recites 61 different kinases however,

claims 182-186 recites 92, 110, 116, 119 and 122 recite purified

active kinases which is broader than the 61 purified kinase

recited in claim 1.

    The text of those sections of Title 35, U.S. Code not

included in this action can be found in a prior Office action.

Claim Rejections - 35 USC § 102/§ 103

    Claims 1-11, 141, 181-186, 188, and 193-195, as amended,

are rejected under 35 U.S.C. 102(a) as anticipated by or, in the

alternative, under 35 U.S.C. 103(a) as obvious over Uetz et al

(Nature, 2/10/2000) for reasons of record as reiterated below.

    Uetz et al, throughout the reference, teach a protein array

representing yeast genome encoded proteins (see Abstract of

the reference). The reference teaches fusing roughly 6000

potential ORFs (genes) from yeast genome (which comprises

approximately 6000 genes) (see page 623, left col., 1st

paragraph. and page 624, left col., 2nd paragraph). Uetz

teaches the yeast proteins were expressed in 96-well assay

plates (page 624, left col., bottom of 2nd paragraph), which

reads on a solid support of the addressable array of claim 1

because each well of the plates would have defined (or

addressable positions). The reference also teach each of the

protein encoded by a gene is expressed individually in

individual wells of the plates as shown in Figure 1 of the

reference (page 624), which reads on each protein being at a

different position on a solid support of claim 1, for example.

The claimed kinase present in the array would have been

inherent to the yeast array taught by Uetz since yeast

inherently contain kinase in its structure or would have been

obvious to determine given the identified genome of yeast as

taught by Uetz.

> Where the claimed and prior art products are identical or
> substantially identical, or are produced by identical or
> substantially identical processes, the PTO can require an
> applicant to prove that the prior art products do not
> necessarily or inherently possess the characteristics of
> his claimed product. See In re Ludtke, supra. Whether the
> rejection is based on "inherency" under 35 USC 102, on
> "prima facie obviousness" under 35 USC 103, jointly or

alternatively, the burden of proof is the same as is evidenced by the PTO's inability to manufacture products or to obtain and compare prior art products. See In re Brown, 59 CCPA 1036, 459 F.2d 531, 173 USPQ 685 (1972); In re Best 195 USPQ 430 (CCPA 1977).

Response to Arguments

Applicants submit that Uetz does not disclose the preparation of an array comprising purified active kinases, and hence, cannot anticipate the presently claimed invention. As set forth in the Methods section of Uetz, at page 627, the disclosed arrays were prepared by transferring patches of transformed yeast cells into wells of a micro-array assay plate. Uetz does not disclose any purification of the yeast proteins prior to placement in the assay plate, just simply transfer of the transformed cells. Hence, Uetz does not disclose the use of purified kinases or functional kinase domains, as recited in present claim 1. Applicants acknowledge that, even assuming the arrays disclosed in Uetz comprise 61 kinases, there is no disclosure in Uetz sufficient to render obvious the construction of an array of at least 61 kinases or functional kinase domains, in which the array comprises kinases that are purified and active, as recited in present claim 1.

In response, applicants' arguments as to the construction of the array are not commensurate in scope with the claims. The claims are drawn to an array and not to a method of making the

array. Nonetheless, attention is drawn to the disclosure of Uetz

at e.g., paragraph bridging col.1 and col. 2 which recites:

> To examine protein activity in a format that allows the
> assay of every predicted ORF: we constructed an array of
> hybrid proteins. At least two general types of protein
> array may be envisioned: those composed of living
> transformants….. and those composed solely of the purified
> proteins (7). The two- hybrid array used here is a set of
> yeast colonies derived from about 6,000 individual
> transformants…...

Claims 1-11, 141, 181-186, 188, and 193-195 are rejected

under 35 U.S.C. 103(a) as being unpatentable over Shalon (WO

95/35505) in view of Felder et al (USP 6458533) or Lafferty(USP

6972183) for reasons of record as repeated below.

Shalon discloses at e.g., page 12, lines 3-9:

> A microarray as an array of regions having a density of
> discrete regions of at least about 100/cm2, and preferably
> at least about 1000/cm2. The regions in a microarray have
> typical dimensions, e.g., diameters, in the range of
> between about 10-250 um, and are separated from other
> regions in the array by about the same distance.

Shalon discloses at e.g., page 30, line 30 up to page 32, line

15:

> Sheets of plastic-backed nitrocellulose where each
> microarray could contain, for example, 100 DNA fragments
> representing all known mutations of a given gene. The
> region of interest from each of the DNA samples from 96
> patients could be amplified, labeled, and hybridized to the
> 96 individual arrays with each assay performed in 100
> microliters of hybridization solution……… In addition to the

genetic applications listed above, arrays of… enzymes….
preparations…..

Shalon discloses an array of enzymes and not kinase as
claimed. However, Feder discloses:

Feder discloses at Example 18:

  Kinases are enzymes that attach a phosphate to proteins.
  Many have been shown to stimulate normal and neoplastic
  cell growth. Hence, compounds that inhibit specific kinases
  (but not all kinases) can be used to test whether the
  kinases are involved in pathology and, if so, to serve as
  starting points for pharmaceutical development… Each kinase
  has substrates that are partially identified, as short
  peptides that contain a tyrosine. Some of the kinase
  specificities overlap so that different kinases may
  phosphorylate some peptides equally but others
  preferentially. For the five kinases, 36 peptide substrates
  are selected that show a spectrum of specific and
  overlapping specificities.

Lafferty discloses at e.g., col. 31, lines 41-49 the
conventionality of an array containing subtrate-enzymes such as
kinase.

Accordingly, it would have been obvious to one having
ordinary skill in the art at the time the invention was made to
use in the array of Shalon the enzyme kinase as taught by Feder.
Feder teaches that kinase have been shown to stimulate normal
and neoplastic cell growth. To use the kinase in the array of
Shalon would lead one having ordinary skill in the art in
determining the kinase in the array responsible for neoplastic

or normal cell growth. Furthermore, as taught by Lafferty an

array containing a kinase is known in the art. [See also

applicants' admission in the response at page 17, of the

12/19/2006 REMARKS. Applicant states: compositions *utilizing*

well-known and well-characterized classes of proteins, as in the

presently claimed invention].

Response to Arguments

Applicants note that Shalon is primarily directed to arrays

comprising polynucleotides (see Examples 1-3), and only

mentions in passing that arrays comprising proteins and

enzymes could be constructed.

In reply, in considering disclosure of a reference, it is

proper to take into account not only specific teachings of

the reference but also "inferences" which one skilled in

the art would reasonably be expected to draw therefrom. In

re Preda 159 USPQ 342. Accordingly, the mention of protein

array in Shalon suffices the prima facie finding of

obviousness.

Felder discloses preparation of arrays comprising

peptides that are substrates for kinases, not arrays

comprising the kinases themselves, "[a] chimeric linker

molecule is prepared in which a 25 base pair

oligonucleotide complementary to one of the anchors is

crosslinked to a peptide substrate of a tyrosine

phosphokinase enzyme." Felder at column 44, lines 18-21

(emphasis added). Thus, Felder does not disclose the

preparation of arrays comprising 61 purified active kinases

or functional kinase domains thereof, as recited in present

claim 1.

With regard to Lafferty, Applicants note that the arrays

disclosed therein are limited to enzymes expressed in

expression library cells, and that Lafferty does not

disclose the purification of these enzymes prior to

placement on a solid support, as recited in the presently

claimed invention.

In response, Felder is employed not for the purpose as

argued rather for its disclosure of the known kinases.

Shalon teaches the arrays of enzymes, to which the specific

enzyme kinase would be prima facie obvious to position

therein as Felder teaches the known kinases. Lafferty is

also employed not for the purpose as argued. Please see the

rejection above. Hence, the combined teachings of the prior

art would lead one having ordinary skill in the art to the

claimed array of purified kinase.

When considering obviousness of a combination of known
elements, the operative question is thus "whether the
improvement is more than the predictable use of prior art

elements according to their established functions." KSR
International Co. v. Teleflex Inc., 550 USPQ2d 1385 (2007).

No claim is allowed.

Any inquiry concerning this communication or earlier
communications from the examiner should be directed to TERESA
WESSENDORF whose telephone number is (571)272-0812. The
examiner can normally be reached on flexitime.
If attempts to reach the examiner by telephone are
unsuccessful, the examiner's supervisor, Christopher Low can be
reached on 571-272-0951. The fax phone number for the
organization where this application or proceeding is assigned is
571-273-8300.
Information regarding the status of an application may be
obtained from the Patent Application Information Retrieval
(PAIR) system. Status information for published applications
may be obtained from either Private PAIR or Public PAIR. Status
information for unpublished applications is available through
Private PAIR only. For more information about the PAIR system,
see http://pair-direct.uspto.gov. Should you have questions on
access to the Private PAIR system, contact the Electronic
Business Center (EBC) at 866-217-9197 (toll-free). If you would
like assistance from a USPTO Customer Service Representative or
access to the automated information system, call 800-786-9199
(IN USA OR CANADA) or 571-272-1000.

/TERESA WESSENDORF/
Primary Examiner, Art Unit 1639

Exhibit 5

UNITED STATES PATENT AND TRADEMARK OFFICE

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
09/849,781	05/04/2001	Michael Snyder	2681.0030002/RWE/JKM	9891

26111   7590   07/09/2010
STERNE, KESSLER, GOLDSTEIN & FOX P.L.L.C.
1100 NEW YORK AVENUE, N.W.
WASHINGTON, DC 20005

EXAMINER
WESSENDORF, TERESA D

ART UNIT	PAPER NUMBER
1639	

MAIL DATE	DELIVERY MODE
07/09/2010	PAPER

Please find below and/or attached an Office communication concerning this application or proceeding.

The time period for reply, if any, is set in the attached communication.

	Application No.	Applicant(s)
Office Action Summary	09/849,781	SNYDER ET AL.
	Examiner	Art Unit	
	TERESA WESSENDORF	1639	

— The MAILING DATE of this communication appears on the cover sheet with the correspondence address —

Period for Reply

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE <u>3</u> MONTH(S) OR THIRTY (30) DAYS, WHICHEVER IS LONGER, FROM THE MAILING DATE OF THIS COMMUNICATION.
- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133).
  Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

Status

1)☒ Responsive to communication(s) filed on <u>07 May 2010</u>.

2a)☒ This action is **FINAL**.   2b)☐ This action is non-final.

3)☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

Disposition of Claims

4)☒ Claim(s) *See Continuation Sheet* is/are pending in the application.

    4a) Of the above claim(s) *See Continuation Sheet* is/are withdrawn from consideration.

5)☐ Claim(s) _____ is/are allowed.

6)☒ Claim(s) *1-11, 141, 164, 169, 173, 174, 177, 181-186, 188 and 193-195* is/are rejected.

7)☐ Claim(s) _____ is/are objected to.

8)☐ Claim(s) _____ are subject to restriction and/or election requirement.

Application Papers

9)☐ The specification is objected to by the Examiner.

10)☐ The drawing(s) filed on _____ is/are: a)☐ accepted or b)☐ objected to by the Examiner.

    Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).

    Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).

11)☐ The oath or declaration is objected to by the Examiner. Note the attached Office Action or form PTO-152.

Priority under 35 U.S.C. § 119

12)☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).

    a)☐ All b)☐ Some * c)☐ None of:

      1.☐ Certified copies of the priority documents have been received.

      2.☐ Certified copies of the priority documents have been received in Application No. _____.

      3.☐ Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).

    * See the attached detailed Office action for a list of the certified copies not received.

Attachment(s)

1)☐ Notice of References Cited (PTO-892)
2)☐ Notice of Draftsperson's Patent Drawing Review (PTO-948)
3)☐ Information Disclosure Statement(s) (PTO/SB/08)
    Paper No(s)/Mail Date _____.

4)☒ Interview Summary (PTO-413)
    Paper No(s)/Mail Date. _____ .
5)☐ Notice of Informal Patent Application
6)☐ Other: _____.

DETAILED ACTION

Status of Claims

Claims 1-16, 93-101, 106, 107, 112-133, 138-159, 162, 164-165, 167, 169, 171, 173-175, 177, 181-186, 188, and 193-197 are pending.

Claims 12-16, 93-101, 106, 107, 112-133, 138-140, 142-159, 162, 165, 167, 171, 175 and 196-197 are withdrawn from further consideration pursuant to 37 CFR 1.142(b), as being drawn to nonelected species, there being no allowable generic or linking claim.

Claims 17-92, 102-105, 108-111, 134-137, 160-161, 163, 166, 168, 170, 172, 176, 178-180, 187, 189-192 and 198 have been cancelled.

Claims 1-11, 141, 164, 169, 173, 174, 177, 181-186, 188 and 193-195 are under consideration in this Office Action.

Withdrawn Objection and Rejections

In view of applicants' arguments and amendments to the claims, the 35 USC 112, second paragraph rejections have been withdrawn.

The text of those sections of Title 35, U.S. Code not
included in this action can be found in a prior Office action.

Claim Rejections - 35 USC § 112

Claims 1-11, 141, 164, 169, 173, 174, 177, 181-186, 188 and
193-195, as amended, are rejected under 35 U.S.C. 112, first
paragraph, as failing to comply with the written description
requirement. The claim(s) contains subject matter, which was
not described in the specification in such a way as to
reasonably convey to one skilled in the relevant art that the
inventor(s), at the time the application was filed, had
possession of the claimed invention for reasons of record as
reiterated below.

Written Description Rejection

A "written description of an invention involving a
chemical genus, like a description of a chemical species,
requires a precise definition, such as by structure, formula
[or] chemical name of the claimed subject matter sufficient to
distinguish it from other materials". University of California

v. Eli Lilly and Col, 43 USPQ 2d 1398, 1405(1997), quoting

Fiefs V. Revel, 25 USPQ 2d 1601m 16106 (Fed. Cir. 1993.

The claimed invention is drawn to a positionally

addressable array comprising a plurality of different

substances on a solid support, with each different substance

being at a different position on the solid support, wherein

the density of the different substances on the solid support

is at least 100 different substances per cm2, and wherein the

plurality of different substances comprises 61 purified active

kinases or functional kinase domains thereof of a mammal, 61

purified active kinases or functional kinase domains thereof

of a yeast, or 61 purified active kinases or functional kinase

domains thereof of a Drosophila.

The specification fails to provide an adequate written

description of 61 purified kinase and functional domains thereof

from any organism such as mammals. The specification provides

general statements of these various kinases in an organism which

are not a detail description of the invention. The detail

description in the specification (Example I, page 27) describes

the 122 kinase genes specifically from the yeast genome, not

from the broad claimed any kinase from any type of mammals or

Drosophila or functional domains thereof. A written description

of a single species would not be a written description for the
genus as claimed. At the time of applicants' invention kinases
in any mammals included in the huge scope of the claim has not
been fully characterized such that it has been positioned in an
array without denaturing the purified protein. A skilled artisan
recognizes that one cannot rule out the possibility that kinases
other than the desired enzyme can contaminate any type of
purification preparations. Notwithstanding this, the kind/type
and preparation of a substrate compatible with the purified
protein is also a factor to consider for a purified protein to
be active in an array. Furthermore, "although most of the
kinases were active in [our] assays, several were not.
Presumably, our preparations of these latter kinases either lack
sufficient quantities of an activator or were not purified under
activating conditions. For example, Cdc28 which was not active
in [our] assays, might be lacking its activating cyclins. For
the case of Hogl, cells were treated with high salt to activate
the enzyme…." (paragraph [0161] of the instant specification,
publication no. 20030207467).

Attention is also drawn to the numerous prior art cited by
applicants, inter alia, the Anderson reference, which teaches
the numerous unforeseeable factors of a purified kinase
positioning in an array.

Anderson states that:

...protein microarrays have still not found widespread use,
in part because producing them is challenging.
Historically, it has required the high-throughput
production and purification of protein, which then must be
spotted on the arrays. Once printed, concerns remain about
the shelf life of proteins on the arrays.

Shaw et al (Drug Discovery and Development, Exhibit B)

concords with the statement that:

"[i]t was first thought that protein biochips would just be
an extension of DNA microarrays, and that hasn't exactly
panned out," says Bodovitz. That's because proteins have
proven to be much trickier to work with in array format
than their genomic counterparts. First of all, there are
issues of stability. Membrane proteins, for example, make
up the majority of potential drug targets, but they're
particularly challenging to stabilize. Then there's the
choice of immobilization technique, which determines how
well the target protein presents itself to the capture
agent, and the problem of nonspecific binding. And of
course, proteins are inherently unstable outside their
natural habitat of living cells, making them much more
challenging than DNA to tag and manipulate.

An applicant shows possession of the claimed invention by

describing the claimed invention with all of its limitations

using such descriptive means as words, structures and formulas

to show that the invention is complete. Lockwood v. American

Airlines, Inc., 107 F.3d 1565, 1572, 41 USPQM 1961, 1966 (Fed.

Cir. 1997); MPEP 2163. Herein, kinase has been described only in

words. The characterization of the different kinases from one

organism to another from the numerous kinases and numerous

organisms has not been adequately described to distinguish one

from the other. To date only a few organisms are fully

characterized and the kinase region has not been fingerprinted

in a partly or even fully characterized gene. The description

lacks structural characterization of a purified kinase as

generically claim. It does not distinguish one kinase from

another and/or one organism from another positioned in any

kind/type of substrate array and reasonably expect the purified

kinase to retain its active form.

Response to Arguments

Applicants submit that kinases and functional kinase

domains from yeast, mammals and Drosophila were a well

characterized group of proteins that were generally known,

understood to be well conserved in structure and function,

easily identified, and readily prepared and assayed by those of

ordinary skill in the art on the priority date of the present

application. Thus, such proteins were well known to those of

ordinary skill in the art, and hence, a re-description of such

proteins is not required under Capon.

In reply, the claims are not drawn only to the alleged

known, well characterized kinases from yeast, mammals and

Drosophila or functional domain thereof. Rather, to a kinase

array for all or any kind of kinase from e.g., mammals or

functional domain immobilized in every conceivable manner on any

kind of solid support. If an appellant choose to rely upon

general knowledge in the art to render his disclosure enabling,

the appellant must show that anyone skilled in the art would

have actually possessed the knowledge, In re Lange (CCPA 1981)

644 F2d 856, 209 USPQ 288, or would reasonably be expected to

check the source which appellant relies upon to complete his

disclosure and would be able to locate the information with no

more than reasonable intelligence. Herein, there is no explicit

description of the generic claim array containing kinases from

any kind of e.g., mammals (except for humans as per the newly

submitted Schweitzer declaration), let alone to the functional

domain thereof. Claims drawn to the use of known chemical

compounds must have a corresponding written description only so

specific as to lead one to that class of compounds. In re

Herschler (CCPA 1979) 200 USPQ 711.

Applicants submit that the level of skill and knowledge

relating to protein kinases and their functional domains was

very high on the priority date of the present application and a

person of ordinary skill in the art would readily understand

that indeed, Applicants were clearly in possession of a

positionally addressable array comprising 61 purified active

kinases or functional kinase domains thereof of a mammal, yeast

or Drosophila.

In reply, the level of skill and knowledge in the art is

high so also the unpredictability in the gene art. This is

demonstrated by no less than applicants for the very specific

yeast ORF genes containing kinase, not the functional domain

thereof. Applicants state at e.g., page 32, lines 14-17:

> "......14 of 15 119 GST::kinase samples were not detected
> by immunoblotting analysis. Presumably, these proteins
> are **not stably overproduced** in the pep4 protease-
> deficient strain used, or **these proteins may form
> insoluble aggregates that do not purify using our
> procedures........"

Please see also the various prior art concording with

applicants' findings above. For example, Anderson states:

> ...protein microarrays have still not found widespread use,
> in part because producing them is challenging.
> Historically, it has required the high-throughput
> production and purification of protein, which then must be
> spotted on the arrays. Once printed, concerns remain about
> the shelf life of proteins on the arrays.

Applicants state that the specification describes the use of positionally addressable arrays containing proteins and functional domains of the proteins from organisms including mammals, yeast and Drosophila (published [0058]); and provides a working example describing the production of a protein chip containing over 100 functional yeast kinases and yeast kinase domains (See Example I).

In reply, the detail description of the yeast array as described in Example I is not controverted. The issue is e.g., an array containing not only yeast but any type of mammal kinase and/or Drosophila (sequence or non-sequence) (and at least 61 kinase in an array).

There is nothing in the description or any prior art teachings of an array containing an immobilized kinase from yeast, mammals and Drosophila or functional domain thereof. It does not describe that the 61 kinase present in yeast can be extrapolated or are similarly present to the different numbers and kinds of kinases found in any kind of mammals or Drosphila.

Attention is again directed to the different prior art cited above as to the high unpredictability in the art for an array containing protein such as kinase.

Claim Rejections - 35 USC § 112

Enablement Rejection

Claims 1-11, 141, 164, 169, 173, 174, 177, 181-186, 188 and

193-195, as amended, are rejected under 35 U.S.C. 112, first

paragraph, because the specification, while being enabling for

yeast protein' kinases of the Ser/Thr and tyrosine kinase

family, does not reasonably provide enablement for the broad

scope of an array of 61 kinases and functional domain kinase

from an organism as mammal, yeast or Drosophila. The

specification does not enable any person skilled in the art to

which it pertains, or with which it is most nearly connected, to

make and use the invention commensurate in scope with these

claims for reasons as repeated below.

The claimed array comprises a broad genus of compositions.

The claimed different substances encompass any members of the

protein kinase from the organism of mammal, yeast, or Drosophila

which is broader than the enabling disclosure. The claimed array

represents enormous scope because the claims do not place any

limitations on the kind, number and/or length of kinase either

singly from one family of organism or a combination(s) from the

different numerous recited organisms. The instant specification

is directed to an array comprising a plurality of different

yeast protein kinase, specifically 122 different yeast protein'

kinases of the Ser/Thr and tyrosine kinase family members (see

specification: example I, pg. 27, line 19 thru pg. 35, line 20;

example II, pg. 41, line 19 thru pg. 43, line 6). The

specification does not provide reasonable assurance to one

skilled in the art that the 61 kinases found in the yeast could

be found in any or all of the organisms such as mammals

especially the functional domain thereof. It is not apparent

from the specification whether the same number of kinases or the

kind of kinases or functional domain thereof can be found in any

other organisms and made into an array. It is not apparent from

the disclosure as to the functional domain of the kinase and the

specific function attributed to said kinase positioned on the

array. The general knowledge and level of skill in the art do

not supplement the omitted description because specific, not

general, guidance is what is needed. In a highly unpredictable

art, as biotechnology, where one cannot predict whether one

species would be predictive to the huge scope of the claim, one

cannot make a priori statement without any experimental studies.

Factors such as the compatibility of the array with the

substrate and compounds disposed therein, the compounds

(kinases) itself and other unpredictable variables can affect

the active form of any kinase. Thus, one cannot predict from a

single species its correspondence or extrapolation to the genus,

as claimed.

Response to Arguments

Applicants cite the Schweitzer Declaration at pages 3-4,

section 8, the Snyder Declaration and Replies to the Office

Action filed December 21, 2007, and April 20, 2009 to support

their enablement position. It is asserted that the replies and

declaration show that protein kinases and functional kinase

domains used in the claimed positionally addressable arrays were

at the time this application was filed, all well-known, and

well-characterized. Reference was made to the Hunter and Plowman

reference.

In reply, as stated above the claims are drawn to an array

of the well-known and well-characterized kinases in yeast and

etc. and not solely to the well-known and characterized yeast,

per se.

Hunter like the specification is drawn only to kinase in

yeast, which is not present in an array. Mammals or Drosophila

has not been described by Hunter or taught that the well known

kinase from yeast, let alone, the functional domain, applies to

other kinases as in mammals or Droosophila.

Applicants further assert that it was also well known at

the time of filing of this application that kinases are highly

conserved such that homologs exist between yeast and many other

organisms. See Manning et al., "The Protein Kinase Complement of

the Human Genome," Science 298:1912-1934 (2002) at page 1913,

first column, first paragraph (cited in Applicants' 6th SIDS

submitted on April 20, 2009). Furthermore, the regulation of the

different kinases and the phosphorylation motifs of substrates

recognized by related kinases are often the same, indicating

that they behave similarly biochemically. See id. Moreover, as

the structure and function of kinases were known to be highly

conserved, it was also known that human kinases can be

substituted for yeast kinases, illustrating the highly conserved

structure-function relationships known to exist for kinases on

the priority date of the application. See Lee and Nurse,

"Complementation used to clone a human homologue of the fission

yeast cell cycle control gene cdc2," Nature 327:31-35 (1987)

(cited in Applicants' 6th SIDS submitted on April 20, 2009).

Therefore, on the priority date of the present invention, the

state of the art relating to protein kinases was extremely high

and was such that a person of skill in the art, in the fields of

for example, protein purification, proteomics and analysis,

enlightened by the teaching of the specification would have

appreciated that no more than routine experimentation would be

required to make and use the claimed arrays containing purified

active kinases or functional kinase domains from a mammal, yeast

or Drosophila.

    In reply, the arguments are not drawn again to kinase in

yeast per se as well as to its properties or homologs thereof.

Rather, the claim is to a kinase of yeast in an array. As Shaw

et al stated above:

> "[i]t was first thought that protein biochips would just be
> an extension of DNA microarrays, and that hasn't exactly
> panned out," says Bodovitz. That's because proteins have
> proven to be much trickier to work with in array format
> than their genomic counterparts. First of all, there are
> issues of stability. Membrane **proteins,** for example, make
> up the majority of potential drug targets, but they're
> **particularly challenging to stabilize.** Then there's the
> **choice of immobilization technique,** which determines how
> well the target protein presents itself to the capture
> agent, and the **problem of nonspecific binding.** And of
> course, proteins are inherently unstable outside their
> natural habitat of living cells, making them much more
> challenging than DNA to tag and manipulate. (Emphasis
> added.)

While the kinases are alleged to be homologs the fact

remains that purification the technique and other experimental

conditions/steps for yeast would be different from any type of

mammals. See applicants' disclosure at e.g., page 32, lines 14-

17 as to the unpredictable or unexpected failure of obtaining

purified yeast in the ORF region alone (i.e., not to its alleged

homologs):

> "……14 of 15 119 GST::kinase samples were not detected
> by immunoblotting analysis. Presumably, these proteins
> are **not stably overproduced** in the pep4 protease-
> deficient strain used, or **these proteins may form
> insoluble aggregates that do not purify using our
> procedures……..**" (Emphasis added.)

Applicants submit that methods useful for confirming kinase

activity of the proteins on the claimed arrays are described in

the specification and were otherwise well known as of the filing

date of the present application (see e.g., Example 1 of

specification). See also Snyder Declaration at pages 3-4,

section 7. Thus protein kinases, functional kinase domains and

methods of assaying these proteins, were well-known in the art

on the priority date of the present invention.

In reply, Example I states that the tyrosine kinase family members do not exist although seven protein kinases that phophorylate have been reported. Applicants' arguments that array from any organisms are simple and straightforward are mere arguments, absent evidence to the contrary, which cannot be substituted for enabling disclosure.

Applicants state for enablement, a specification need not teach, and preferably omits, information that is well-known to those of ordinary skill in the art. See Hybritech Inc. v. Monoclonal Antibodies, Inc., 802 F.2d 1367, 1384 (Fed. Cir. 1986); Lindemann Maschinetzfabrik v. American Hoist and Derrick, 730 F.2d 1452, 1463 (Fed. Cir. 1984); In re Wands, 8 USPQ2d 1400, 1402 (Fed. Cir. 1988).

In reply, the Federal Circuit has cautioned against over reliance on the assertion that everything needed to practice the full scope of the claims was "known in the art" and that a patent need not teach, and preferably omits, what is well known in the art. See Genentech Inc. v. NovoNordiskA/S, 108 F.3d 1361, 1366, 42 USPQ2d 1001, 1005 (Fed. Cir. 1997): "[T]hat general, oft-repeated statement is merely a rule of supplementation, not a substitute for a basic enabling disclosure. It means that the

omission of minor details does not cause a specification to fail

to meet the enablement requirement .... It is the specification,

not the knowledge of one skilled in the art that must supply the

novel aspects of an invention in order to constitute adequate

enablement." Herein the specification teaches the kinase of

yeast only in the ORF region (not even of the full length yeast

sequence. Do the kinase and its functional domain exist only in

the ORF region and not in any other region(s) of the yeast

gene?) Applicants point to nothing in the specification that

would indicate to the contrary (i.e., kinase array of e.g., any

mammal or Drosophila in any region of the full length sequence

or unsequence protein).

Applicants rely upon the newly submitted Schweitzer

Declaration, in addition to the Snyder declaration. Schweitzer

is stated to describe the preparation of functional human

protein kinase arrays using the teaching in the present

specification. In addition, the functional human kinase domains

used in the positionally addressable arrays prepared by

Schweitzer that form the basis of the present claims were, on

the priority date of the present application, well-known, well-

characterized proteins with purified human kinases (see the

Schweitzer Declaration, at page 3, section 8). As discussed in

detail in the Schweitzer Declaration at pages 4-7, sections 9-

13, researchers enlightened by the information set forth in the

specification, have used the homologies that were known to exist

between human and yeast kinases, to informatically identify

genes for human kinases and functional domains, clone these

genes, express these genes in Sf9 insect cells, lyse the cells

and purify the human kinases and functional domains. (See also

Protein-Protein Interaction Profiling on Invitrogen ProtoArray

TM High-Density Protein Microarrays, Application Note,

Invitrogen page 2, column 2, paragraph 3 - 2, column 3,

paragraph 1 (2005) (hereinafter "Protein-Protein Interaction

Profiling," Exhibit B)). According to the Schweitzer

Declaration, over 90% of protein kinases expressed and purified

using the methods described in the specification were active as

demonstrated by catalytic activity including

autophosphorylation, wherein a protein kinase phosphorylates

itself. See id.

In reply, that the Schweitzer declaration teaches the

purification of kinase from humans is not controverted. The

Schweitzer declaration describes kinase allegedly obtained from

human not from any mammals or Drosophila. It is not clear from

the Schweitzer declaration just which part of the present

disclosure has been applied from the yeast to the human kinase

array, except to get its homology therefrom. The instant

specification uses chip for its yeast array. Schweitzer does not

teach a biochip.

Exhibit B presents a description for the human array which

does not seem to fall or correspond to the description for

yeast. Furthermore Exhibit C of the Schweitzer declaration

states at page 1:

> The family of human protein kinases consists of more than
> 500 members of which only a fraction have been
> characterized to date. Much is still not known about the
> biological function of many kinases, the protein substrates
> that are phosphorylated by these kinases, or the roles of
> these kinases and substrates in disease..

Thus, Schweitzer has not extrapolated or predicted its

findings to any other family o human protein kinases which

consist of more than 500 members to which only a fraction has

been characterized to date.

Claim Rejections - 35 USC § 102/§ 103

Claims 1-11, 141, 181-186, 188, and 193-195, as amended,

are rejected under 35 U.S.C. 102(a) as anticipated by or, in the

alternative, under 35 U.S.C. 103(a) as obvious over Uetz et al

(Nature, 2/10/2000) for reasons of record as reiterated below.

Uetz et al, throughout the reference, teach a protein array

representing yeast genome encoded proteins (see Abstract of

the reference). The reference teaches fusing roughly 6000

potential ORFs (genes) from yeast genome (which comprises

approximately 6000 genes) (see page 623, left col., 1st

paragraph. and page 624, left col., 2nd paragraph). Uetz

teaches the yeast proteins were expressed in 96-well assay

plates (page 624, left col., bottom of 2nd paragraph), which

reads on a solid support of the addressable array of claim 1

because each well of the plates would have defined (or

addressable positions). The reference also teach each of the

protein encoded by a gene is expressed individually in

individual wells of the plates as shown in Figure 1 of the

reference (page 624), which reads on each protein being at a

different position on a solid support of claim 1, for example.

The claimed kinase present in the array would have been

inherent to the yeast array taught by Uetz since yeast

inherently contain kinase in its structure or would have been

obvious to determine given the identified genome of yeast as

taught by Uetz.

Where the claimed and prior art products are identical or
substantially identical, or are produced by identical or
substantially identical processes, the PTO can require an
applicant to prove that the prior art products do not
necessarily or inherently possess the characteristics of
his claimed product. See In re Ludtke, supra. Whether the
rejection is based on "inherency" under 35 USC 102, on
"prima facie obviousness" under 35 USC 103, jointly or
alternatively, the burden of proof is the same as is
evidenced by the PTO's inability to manufacture products or
to obtain and compare prior art products. See In re Brown,
59 CCPA 1036, 459 F.2d 531, 173 USPQ 685 (1972); In re
Best 195 USPQ 430 (CCPA 1977).

Response to Arguments

Applicants state that Uetz does not disclose the claim

arrays comprising purified kinases or functional kinase domains.

Applicants assert that Uetz pg 623, col. 1, last paragraph –

col. 2, first paragraph provides further evidence that the

arrays in Uetz did not consist of purified proteins having

kinase activity.

In reply, attention is again drawn to the disclosure of

Uetz at e.g., paragraph bridging col.1 and col. 2 which recites

a purified yeast ORF (referring to reference 7) which contains

the kinase region):

> To examine protein activity in a format that
> allows the assay of every predicted ORF: we
> constructed an array of hybrid proteins. At least two
> general types of protein array may be envisioned:

> those composed of living transformants….. and those
> composed solely of **the purified proteins (7)**. The two-
> hybrid array used here is a set of yeast colonies
> derived from about 6,000 individual transformants…..
> (Emphasis added.)

Thus, Uetz discloses purified protein containing the same

kinase from yeast. Even assuming that Uetz does not disclose

(but Uetz does) a purified kinase as argued however, the

unpurified ORF of the yeast containing kinase would be the same

as the claimed purified one. A purified kinase obtained from the

same source as yeast merely further characterizes the known

kinase present in the yeast. Applicants' use of the word

comprising does not preclude the other elements present in the

kinase contain in the ORF region of the yeast. As applicants

stated above:

> Kinases and functional kinase domains from yeast, mammals
> and Drosophila were a well characterized group of proteins
> that were generally known, understood to be well conserved
> in structure and function, easily identified, and readily
> prepared and assayed by those of ordinary skill in the art
> are known.

Applicants note that, even assuming the arrays disclosed

in Uetz comprise 61 kinases, there is no disclosure in Uetz

sufficient to render obvious the construction of an array of 61

kinases or functional kinase domains, in which the array

comprises kinases that are purified and active, as recited in

present claim 1.

Applicants agree that the claims are drawn to an array but

assert that the limitation purified and active recited in the

claims is not a process limitation, but rather a characteristic

of the components of the array.

In reply, as recognized by applicants, a purified kinase is

but a further characteristics of the known (albeit, allegedly

unpurified) compound. Thus, this further characterization does

not make the compound, kinase any different but only further

characterizes said known kinase. It is well settled that there

can be no patentable invention where novelty does not exist,

albeit all of the properties of said compositions were not

previously recognized. (Please see further the above statements

of applicants that these kinases are known).

Likewise, it is well settled in the art that where

substance having medicinal properties is produced, it becomes an

immediate consideration to prepare substances in as pure a form

as possible. Claim for known substance which differs from prior

art only in degree, as for example in purity, is not patentable.

See Ex parte Steelmand and Kelly, 140 USPQ 189.

Most of applicants'' subsequent arguments rely on the responses filed on December 21, 2007 and the April 20, 2009 replies and the Snyder declaration. Applicants state that the following exemplary references describe the skepticism from those in the field regarding the **preparation** of protein arrays both before and after the time of filing of the present application, as well as some of the problems regarding **preparation of protein arrays** comprising large numbers of purified active proteins that **were overcome by the presently claimed invention**"); Schweitzer declarations and the reference to e.g., Anderson, all relate to the problems encountered into the making of the array and how the problems have been overcome. (Emphasis added.) However, as stated above and in the previous Office actions the claims are drawn to known kinases the immobilization thereof into a solid surface is well known in the art as taught by Uetz above.

Applicants state that the examiner appears to agree that kinases of the various organisms such as mammals, Drosophila and yeast were well known at the time of filing the application. This is in stark contrast to the Examiner's contrary position noted above with regard to written description and enablement of the presently claimed invention. While Applicants agree that

kinases of Drosophila, yeast and mammals were well known in the
art at the time of filing the present application, it would not
have been obvious to place these kinases on a positionally
addressable array so that they were not only purified, but also
active.

In reply, no contradiction exists in the rejections under
obviousness and enablement/written description. These are two
separate rejections. The enablement/written description are
based on the lack of description for the broad claim genus and
not to the yeast species containing kinase. This species is
taught by the prior art which is included in the broad genus
claim hence, anticipating or rendering obvious the broad genus
claim. Applicants' further arguments relying on the Synder
declaration as to method of making the array and Bussow are as
stated above not commensurate with the claims, which recite
simply the known kinases immobilized on solid support.

Claims 1-11, 141, 181-186, 188, and 193-195 are rejected
under 35 U.S.C. 103(a) as being unpatentable over Shalon (WO
95/35505) in view of Felder et al (USP 6458533) or Lafferty(USP
6972183) for reasons of record as repeated below.

Shalon discloses at e.g., page 12, lines 3-9:

> A microarray as an array of regions having a density of
> discrete regions of at least about 100/cm2, and preferably
> at least about 1000/cm2. The regions in a microarray have
> typical dimensions, e.g., diameters, in the range of
> between about 10-250 um, and are separated from other
> regions in the array by about the same distance.

Shalon discloses at e.g., page 30, line 30 up to page 32, line

15:

> Sheets of plastic-backed nitrocellulose where each
> microarray could contain, for example, 100 DNA fragments
> representing all known mutations of a given gene. The
> region of interest from each of the DNA samples from 96
> patients could be amplified, labeled, and hybridized to the
> 96 individual arrays with each assay performed in 100
> microliters of hybridization solution......... In addition to the
> genetic applications listed above, arrays of... enzymes...(were
> prepared].

Shalon discloses an array of enzymes and not kinase as

claimed. However, Feder discloses:

Feder discloses at Example 18:

> Kinases are enzymes that attach a phosphate to proteins.
> Many have been shown to stimulate normal and neoplastic
> cell growth. Hence, compounds that inhibit specific kinases
> (but not all kinases) can be used to test whether the
> kinases are involved in pathology and, if so, to serve as
> starting points for pharmaceutical development... Each kinase
> has substrates that are partially identified, as short
> peptides that contain a tyrosine. Some of the kinase
> specificities overlap so that different kinases may
> phosphorylate some peptides equally but others
> preferentially. For the five kinases, 36 peptide substrates
> are selected that show a spectrum of specific and
> overlapping specificities.

Lafferty discloses at e.g., col. 31, lines 41-49 the conventionality of an array containing substrate-enzymes such as kinase.

Accordingly, it would have been obvious to one having ordinary skill in the art at the time the invention was made to use in the array of Shalon the enzyme kinase as taught by Feder. Feder teaches that kinase have been shown to stimulate normal and neoplastic cell growth. To use the kinase in the array of Shalon would lead one having ordinary skill in the art in determining the kinase in the array responsible for neoplastic or normal cell growth. Furthermore, as taught by Lafferty an array containing a kinase is known in the art. [See also applicants' admission in the response at page 17, of the 12/19/2006 REMARKS. Applicant states: compositions *utilizing well-known and well-characterized classes of proteins*, as in the presently claimed invention].

Response to Arguments

Applicants argue that Shalon is primarily directed to arrays comprising polynucleotides (see Examples 1-3), and only mentions in passing that arrays comprising proteins and enzymes

could be constructed. Furthermore, Felder discloses preparation

of arrays comprising peptides that are substrates for kinases,

not arrays comprising the kinases themselves. Thus, Felder does

not disclose the preparation of arrays comprising 61 purified

active kinases or functional kinase domains thereof, as recited

in present claim 1. With regard to Lafferty, Applicants note

that the arrays disclosed therein are limited to enzymes

expressed in expression library cells, and that Lafferty does

not disclose the purification of these enzymes prior to

placement on a solid support, as recited in the presently

claimed invention. Applicants rely upon the Synder declaration

as support that Lafferty does not disclose the purified enzyme

on a solid support. Lafferty is alleged to use an impure clone

of an enzyme repeatedly passed through a capillary array several

times.

    In reply, much of applicants' arguments are drawn to the

method of making the array, which is not commensurate in scope

with the claims. Shalon,as recognized by applicants above, only

mentions in passing arrays. However, this "passing remarks" or

inferential teachings suffice the findings of obviousness. In

considering disclosure of a reference, it is proper to take into

account not only specific teachings of the reference but also

"inferences" which one skilled in the art would reasonably be expected to draw therefrom. In re Preda 159 USPQ 342.

The disclosure of Lafferty as discussed in the Synder declaration of passing the clone several times into a capillary array and producing an optically detectable signal would indicate a purified product to enable detection of the clone. The claimed number of 61 kinase is dependent upon the organism and location from which the kinase is contained. Thus, this number may be arbitrary considering that the kinase is only derived from the ORF region of yeast. To determine the number of kinases in an organism as yeast in a specific location would be within the ordinary skill in the art, as evidenced from the various well known kinases in a protein sequence.

Applicants cannot attack the references individually when the rejection is based on combination of references. Felder and Lafferty are employed for its disclosure of purified kinase, as claimed, not that it has to teach purifying the kinase prior to immobilization, otherwise it would be anticipatory rejection. Shalon teaches ORF containing kinase on an array but does not expressly teach, albeit implicitly, the purified kinase hence, the application of the secondary references Felder and Lafferty that renders the claim prima facie obvious. It would be within one having ordinary skill in the art at the time the invention

was made to position a known compound as kinase into an array,
as taught by Shalon. There is nothing new and unobvious in mere
positioning a known compound in e.g., array, when in nature
these kinases are inherently arrayed or attached to e.g., a
membrane, which would read on a substrate of an array.

Applicants assert that it was unexpected that kinases and
functional kinase domains of these kinases could be purified and
placed on a solid support to form an array, and that these
kinases and kinase domains would retain their/kinase activity.
As detailed in the Snyder Declaration and the Schweitzer
Declaration, it is only after the guidance provided in the
present specification that a person of ordinary skill in the art
would consider it possible to generate the presently claimed
arrays. As discussed above and in the Schweitzer Declaration,
Applicants respectfully submit that at the time of filing of the
present application, it was unexpected that kinases and
functional kinase domains of these kinases could be purified and
placed on a solid support to form an array, it was also
unexpected that the purified kinases and functional kinase
domains of these kinases would retain their activity when placed
onto the array.

In reply, positioning of a known compound, be it purified
or not, would be expected since the prior art, Shalon has
successfully applied an enzyme protein in an array. There is
nothing novel or unobvious of mere positioning or attaching a
known compound, as kinase, as admitted by applicants above, in
an array. The numerous advantages derived in arraying a known
compound e.g., high throughput screening would provide
motivation to one having ordinary skill in the art at the time
of filing. One would have a reasonable expectation of success in
immobilizing the yeast ORF containing kinase in an array as
successfully made by Shalon and others in the prior art.

[Applicants' arguments above are mostly drawn to what
appears to be the method of making an array immobilized
with purified kinase. Perhaps, this might very well be
where the novelty resides. It is therefore suggested that
applicants draft/amend the claims to recite a method of
making/using the array. The method claim may be an
allowable subject mater in view of the alleged and argued
unexpected results of the method of purifying and attaching
kinase to a solid support.]

        No claim is allowed.

                              Conclusion

        THIS ACTION IS MADE FINAL. Applicant is reminded of the
extension of time policy as set forth in 37 CFR 1.136(a).
        A shortened statutory period for reply to this final action
is set to expire THREE MONTHS from the mailing date of this
action. In the event a first reply is filed within TWO MONTHS
of the mailing date of this final action and the advisory action
is not mailed until after the end of the THREE-MONTH shortened
statutory period, then the shortened statutory period will
expire on the date the advisory action is mailed, and any
extension fee pursuant to 37 CFR 1.136(a) will be calculated
from the mailing date of the advisory action. In no event,
however, will the statutory period for reply expire later than
SIX MONTHS from the mailing date of this final action.

        This application contains claims 12-16, 93-101, 106, 107,

112-133, 138-140, 142-159, 162, 165, 167, 171, 175 and 196-197

drawn to a non elected invention. A complete reply to the final

rejection must include cancellation of nonelected claims or

other appropriate action (37 CFR 1.144) See MPEP § 821.01.

        Any inquiry concerning this communication or earlier
communications from the examiner should be directed to TERESA
WESSENDORF whose telephone number is (571)272-0812. The
examiner can normally be reached on flexitime.
        If attempts to reach the examiner by telephone are
unsuccessful, the examiner's supervisor, Christopher Low can be
reached on 571-272-0951. The fax phone number for the
organization where this application or proceeding is assigned is
571-273-8300.

Information regarding the status of an application may be obtained from the Patent Application Information Retrieval (PAIR) system. Status information for published applications may be obtained from either Private PAIR or Public PAIR. Status information for unpublished applications is available through Private PAIR only. For more information about the PAIR system, see http://pair-direct.uspto.gov. Should you have questions on access to the Private PAIR system, contact the Electronic Business Center (EBC) at 866-217-9197 (toll-free). If you would like assistance from a USPTO Customer Service Representative or access to the automated information system, call 800-786-9199 (IN USA OR CANADA) or 571-272-1000.

/TERESA WESSENDORF/
Primary Examiner, Art Unit 1639

Exhibit 6

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re application of:	Confirmation No.: 9891
SNYDER *et al.*	Art Unit: 1639
Application No.: 09/849,781	Examiner: WESSENDORF, Teresa D.
Filed: May 4, 2001	Atty. Docket: 2681.0030002/RWE/JKM
For: **Protein Chips for High Throughput Screening of Protein Activity**	

Declaration of Barry Schweitzer, Ph.D. Under 37 C.F.R. § 1.132

Attn: Mail Stop Amendment

Commissioner for Patents
PO Box 1450
Alexandria, VA 22313-1450

Sir:

The undersigned, Barry Schweitzer, residing at 459 Maple Avenue, Cheshire, CT, USA, declares and states as follows:

1. I am currently employed by Life Technologies Inc. (hereinafter "LTI"), a licensee of the above-captioned application. I hold the positions of Director of Integrated Technologies, Molecular Biology Systems Division. My credentials are provided in the curriculum vitae that is attached to this declaration as Exhibit A. I received my Ph.D. degree in Pharmacology from Yale University. As seen from my attached curriculum vitae, I have published many papers related to protein microarrays. Based on my education and experience, I am an expert in the field of yeast and human genomics, proteomics, and molecular genetics.

2. I have reviewed and am familiar with U.S. Application No. 09/849,781, (hereinafter "the '781 application") filed on May 4, 2001, the Office Action dated July 7, 2009 ("the Office Action"), issued by the U.S. Patent and Trademark Office in the

Claim 188.   (Previously presented) The positionally addressable array of claim 1, wherein the kinases are yeast kinases.

Claims 189-192.   (Canceled).

Claim 193.   (Previously presented) The positionally addressable array of claim 1, wherein the different substances are 61 purified active kinases.

Claim 194.   (Currently amended) The positionally addressable array of claim 193, wherein the kinases are ~~members-of-the~~ serine/threonine kinase family members, ~~members-of-the~~ tyrosine kinase family members, or ~~the kinases are members of the~~ serine/threonine kinase ~~family~~ and ~~members of the~~ tyrosine kinase family members.

Claim 195.   (Currently amended) The positionally addressable array of claim 1, wherein the functional kinase domains are functional kinase domains of ~~members of the~~ serine/threonine kinase family members, functional kinase domains of ~~members of the~~ tyrosine kinase family members, or ~~wherein the~~ functional kinase domains ~~comprise functional kinase domains~~ of ~~kinases that are members of the~~ serine/threonine kinase family members and functional kinase domains of ~~kinases that are members of the~~ tyrosine kinase family members.

Claim 196.   (Withdrawn) The positionally addressable array of claim 1, wherein the kinases or functional kinase domains are recombinant proteins.

Claim 197.   (Withdrawn) The positionally addressable array of claim 196, wherein the recombinant proteins are recombinant fusion proteins.

Claim 198.   (Canceled).

Atty. Dkt. No. 2681.0030002/RWE/KKH

present application, and the currently pending claims, filed in the Reply to Office Action with this declaration.

3.   The '781 application presently claims a positionally addressable array comprising a plurality of different substances on a solid support, with each different substance being at a different position on the solid support, wherein the density of the different substances on the solid support is at least 100 different substances per cm^2, and wherein the plurality of different substances comprises 61 purified active kinases or functional kinase domains thereof of a mammal, 61 purified active kinases or functional kinase domains thereof of a yeast, or 61 purified active kinases or functional kinase domains thereof of a Drosophila.

4.   In the Office Action at pages 7-12, the Examiner asserts that the claims are allegedly not enabled. Specifically, the Examiner asserts that the specification does not enable the claimed array comprising 61 kinases and functional kinase domains from a mammal, yeast or Drosophila.

5.   In making this declaration, it is my opinion that at the time this application was filed, an ordinary practitioner in the field of genomics and proteomics would have been able to make and use the presently claimed positionally addressable arrays based on knowledge available to those in the field in combination with the detailed disclosure of the '781 application. It is also my opinion that any experimentation required for making and using the presently claimed positionally addressable arrays would have been routine and thus not inordinate or excessive.

6.   As discussed in detail below, the specification of the '781 application

clearly provides sufficient disclosure for a typical practitioner in the field of proteomics

to make and use positionally addressable arrays comprising 61 purified active kinases or

functional kinase domains thereof of a mammal, 61 purified active kinases or functional

kinase domains thereof of a yeast, or 61 purified active kinases or functional kinase

domains thereof of a Drosophila. Based upon this disclosure, in combination with what

was known at the time of filing of the present application, the use of kinases from

various organisms, including mammals and Drosophila, in the preparation of the

presently claimed positionally addressable arrays, would not have required undue

experimentation, but rather, routine and straightforward experiments.

8.   Protein kinases and functional kinase domains used in the positionally

addressable arrays that form the basis of the present claims were, at the time this

application was filed, all well-known, and well-characterized. *See* Hunter and Plowman,

"The protein kinases of budding yeast: six score and more," *TIBS 22*:18-22 (1997) at

page 18, first column, first paragraph (cited in Applicants' 6th SIDS submitted on April

20, 2009). It was also well recognized at the time of filing of this application that

kinases are highly conserved such that homologs exist between yeast and many other

organisms. *See* Manning *et al.*, "The Protein Kinase Complement of the Human

Genome," *Science 298*:1912-1934 (2002) at page 1913, first column, first paragraph

(cited in Applicants' 6th SIDS submitted on April 20, 2009). Furthermore, the regulation

of the different kinases and the phosphorylation motifs of substrates recognized by

related kinases are often the same, indicating that they behave similarly biochemically.

See id. Furthermore, as function is often highly conserved, human kinases can be

substituted for yeast kinases, illustrating the highly conservative nature of these proteins.

See Lee and Nurse, "Complementation used to clone a human homologue of the fission

yeast cell cycle control gene *cdc2*," *Nature 327*: 31-35 (1987) (cited in Applicants' 6th

SIDS submitted on April 20, 2009). Therefore, at the time this application was filed, the

"state of the art" in protein kinases was such that a practitioner possessing a typical level

of skill in proteomics, such skill including but not limited to gene cloning, protein

expression and purification and analysis, would have readily recognized from the '781

application, and the knowledge available in the art, that kinases of yeast, mammals and

Drosophila could routinely be utilized to practice the presently claimed invention.

9.   The Examples set forth in the '781 application describe positionally

addressable protein arrays made with purified, active kinases isolated from yeast. The

specification describes that the purified, active yeast kinases were prepared by cloning

yeast kinase genes into a high copy URA3 expression vector. *See* the '781 application at

page 28, lines 28-30. The plasmids containing the vector sequences were transformed

into yeast, and *Ura*+ colonies were selected. Plasmids were rescued in *E. coli*, then

transformed into the pep4 yeast strain for kinase protein purification. *See id.* at page 28

line 36 to page 29 line 7. Purified, active kinases were attached to polydimethylsiloxane

(PDMS) chips, and the chips comprising the purified, active yeast kinases were assayed

for the phosphorylation of 17 different substrates to determine *in vitro* kinase activity.

See id. at page 33 line 14 to page 34 line 19. *See id.* at page 29 line 26 to page 30

line 21.

10.   In 2000, while working for Molecular Staging, developing antibody based

arrays, I read Dr. Michael Snyder's paper entitled "Analysis of yeast protein kinases

using protein chips." *See* Zhu et al., "Analysis of yeast protein kinases using protein

chips," *Nature Genetics 26*: 283-289 (2000). Dr. Snyder's paper described work that was

extremely impressive and unexpected. Dr. Snyder's discoveries, in fact, motivated me to

accept a position at Protometrix, Branford, CT, a protein array company that had

licensed Dr. Snyder's technology. When I joined Protometrix, work was already

underway to use the information set forth in the '781 application, and known in the art at

the time regarding protein kinases and their highly conserved homology between yeast

and many other organisms, to develop protein arrays with active human kinases.

11.   Using the information set forth in the '781 application and the known

homologies between human and yeast kinases, researchers at Protometrix were able to

informatically identify genes for human kinases and utilize algorithms to identify many

kinase functional domains. The genes were cloned into a recombinant bacmid

(baculovirus shuttle vector), transfected into Sf9 insect cells, and cultured in 96 well

plates. *See* Protein-Protein Interaction Profiling on Invitrogen ProtoArray™ High-

Density Protein Microarrays, Application Note, Invitrogen page 2, column 2, paragraph

3 (2005) (hereinafter "Protein-Protein Interaction Profiling," Exhibit B). In many cases,

efforts were made to clone the full length protein kinases as well as the kinase active

catalytic domains. All the proteins were expressed as N-terminal glutathione-S-

transferase (GST) fusion proteins. *See id.* page 2, column 1, paragraph 1.

12.   After a growth period, the cells were harvested and lysed under

nondenaturing conditions as in the '781 application. *See* Protein-Protein Interaction

Profiling at page 2, column 3, paragraph 1. *See* the '781 application at page 29,

line 8-16. The lysates were further loaded and eluted off of glutathione resin in 96 well

plates under nondenaturing conditions. *See* Protein-Protein Interaction Profiling at page 2, column 3, paragraph 1. Over 90% of protein kinases expressed and purified using the methods described in the '781 application were active as demonstrated by catalytic activity including autophosphoylation, wherein a protein kinase phosphorylates itself. *See id.* In contrast, only approximately 10% of protein kinases were active when other methods known at the time were utilized, including the use of high throughput methods with kinase expression in E. coli.

13.   Finally, using the purified kinases I have described, Protometrix developed a positionally addressable array comprising at least 100 different proteins on a solid support, with each different protein being at a different position on the solid support, wherein the density of the different proteins on the solid support was at least 100 different proteins per cm^2, and contained at least 61 purified active human kinases or functional kinase domains, as presently claimed in the '781 application. Protometrix's technology, sold as Invitrogen's Human ProtoArray High Density Protein Microarrays™, are manufactured with thousands of different quality controlled recombinant human proteins and contain approximately 400 active human kinases and functional kinase domains. *See* B. Schweitzer *et al.*, Development and Validation of Kinase Substrate Screening on Human ProtoArray High Density Protein Microarrays™, Invitrogen, Inc., page 2, column 1, paragraph 2 to page 3, column 1, paragraph 1 (2004) (hereinafter "Schweitzer") (Exhibit C). *See also* Access to the Human Proteome on a Microarray Scale, Invitrogen, Inc., Tables 1 & 2 (2007) (hereinafter "Access to Human Proteome") (Exhibit D). Several commercial versions of this array have been sold with between 1,500 to 9,000 human proteins. The activity of the arrayed kinases has been verified,

including demonstrated catalytic activity by incubating the arrays with radioactive ATP

and measuring autophosphorylation. *See* Schweitzer, page 2, column 2, paragraph 1 to

page 3, column 1, paragraph 1.

14.   It is my opinion that the '781 application provides a clear disclosure of

how to make and use the presently claimed positionally addressable arrays. Those

working in the field of proteomics at the time this application was filed were aware that

yeast, mammalian and Drosophilia kinases are highly conserved and homologous, and

therefore that the teachings of purified, active yeast kinase arrays in the '781 Application

could be used to routinely prepare similar arrays with mammalian and Drosophilia

kinases.   Equipped with this information, researchers at Protometrix, using the

information disclosed in the '781 application, were able to develop protein arrays

comprising human kinases. *See*, e.g., the '781 application at pages 25-35. Thus, in my

opinion, a typical practitioner in the field of proteomics would consider the production of

an array using at least 61 purified, functional kinases from yeast as detailed in the '781

application, to also allow for the routine production and use of arrays comprising

purified, active kinase and functional kinases domains from other organisms, including

mammals and Drosophila, as set forth in the presently claimed invention.

BARRY SCHWEITZER, Ph.D.

CONTACT INFORMATION

459 Maple Avenue
Cheshire, CT 06410
203-439-9439 (Work)
203-645-2953 (Mobile)
barry.schweitzer@lifetech.com

PROFESSIONAL EXPERIENCE

INVITROGEN CORPORATION (Now LIFE TECHNOLOGIES), Carlsbad, CA   2004 - Present
Director, Integrated Technologies, Molecular Biology Systems Division – 2009 to present
Director, Protein Analysis R&D - 2008
Director, Protein Array R&D and Site Leader - 2006 – 2007
Director, Protein Array R&D and Operations - 2004 - 2006
Current responsibilities include the oversight of programs which span the traditional segments of the Molecular Biology Reagent Business, particularly programs that integrate instrumentation with consumables. Previous responsibilities included oversight of R&D and Services for Invitrogen's Protein Analysis product lines, including protein separation technologies, Western technologies, mass spectroscopy, and protein arrays. Additional responsibilities included site leadership of the Protein Array Center in Branford, CT, including R&D, Services, Manufacturing, Quality, and Facilities functions. Other responsibilities include budget preparation and implementation, intellectual property management, and oversight of academic, government, and industrial collaborations and contracts. Also participating in technology and intellectual property evaluations, business development, grant preparations, community relations, and presentations at national and international meetings. Reporting to the Vice President, R&D of the Molecular Biology Reagents Business Unit.

Leadership accomplishments include:
- Led transfer of all operations from Branford, CT to Carlsbad on-time, under budget, and without loss of revenue
- Led global launch of several new multimillion dollar products
- Championed Lean Six Sigma Black Belt and Green Belt projects
- Led ISO 9001 Certification of Branford Site
- Led successful completion of multimillion dollar Biodefense projects in partnership with the United States Army Medical Research Institute for Infectious Diseases (USAMRIID)
- Authored or co-authored 11 publications, including paper in *Nature*
- Inventor or co-inventor on 10 new patent applications
- Presented at 14 international scientific conferences.

PROTOMETRIX, INC., Branford, CT   2002 - 2004
Senior Director, Technology - 2003-2004
Director of Technology – 2002 – 2003
Fifth person to join start-up biotechnology company. Director of a research and development operation providing high-throughput gene cloning, protein expression, protein purification, and protein microarray manufacturing for products, services, and discovery. Additional responsibilities included leading product development teams, leading technology and intellectual property diligence reviews, presenting to investors, coordinating industrial collaborations, and managing prosecution of company intellectual property. Reported to the Vice President, R&D.

Leadership accomplishments included:
- Led the Protometrix technical and IP diligence team during the acquisition of the company by Invitrogen Corp.
- Led the commercial launch of the world's first functional protein microarray product.
- Established the 1st manufacturing facility for the production of protein arrays.
- Built highly skilled team of scientists, engineers, and informatics specialists
- Led the design and buildout of 14,000 s.f. state-of-the-art laboratory and company headquarters.

MOLECULAR STAGING, INC., New Haven, CT   1998 - 2002
Director of Proteomics - 2001 – 2002
Section Head - 1998-2000
Second person to join start-up biotechnology company. Director of a research and service operation providing high-throughput protein expression profiling data using proprietary protein microarray technology to academic, government, and corporate clients. Responsibilities included management of research personnel, budget preparation and implementation, business development, oversight of academic collaborators, preparation of publications and patent applications, presentations for investors, corporate partners and at national meetings. Four direct and 23 indirect reports. Reporting to Chief Operating Officer.

Leadership accomplishments included:
- Successfully launched the world's first microarray-based protein expression profiling service.
- Developed the world's most advanced manufacturing facility for production of antibody microarrays.
- 8 publications, including publication in Nature Biotechnology of 1st application of antibody microarrays for protein expression profiling.
- 1 issued patent, and 4 patent applications.
- Led and coordinated the design and buildout of 46,000 s.f. of state-of-the-art proteomics laboratory.
- Successfully moved an academic technology into an industrial setting, increasing sensitivity, robustness, and utility.
- Led project resulting in $9 MM equity investment by Fortune 100 Company.
- Gave technical presentations resulting in $40 MM 2nd round financing.

WALT DISNEY MEMORIAL CANCER CENTER, Orlando, FL   1994 - 1998
Division Director. Laboratory director of multidisciplinary research program in the structural biology of nucleic acids, proteins, and drugs involved in cancer and related diseases. Responsibilities included carrying out experiments and data analysis, project development, management of 15-20 research, administrative, and volunteer personnel, budget preparation and implementation, grant writing, preparation of publications, public relations, and mentoring of graduate, undergraduate and high school students.
Scientific Director Molecular Diagnostics Clinical Laboratory. Responsibilities included business plan preparation and implementation, management of technical staff, technical consultant, clinical research director, and physician outreach.

Leadership accomplishments included:
- Established and directed a program utilizing multidimensional nuclear magnetic resonance (NMR) spectroscopy, and computational chemistry to determine high-resolution structures of proteins, nucleic acids, and drug complexes for the purpose of chemotherapeutic development.
- Established and directed a laboratory utilizing the most advanced molecular techniques to diagnose infectious diseases, cancer, and inherited diseases for patients of Florida Hospital (2nd largest number of admissions in U.S.).

UNIVERSITY OF CENTRAL FLORIDA, Orlando, FL   1994 - 1998
Assistant Professor
Responsibilities included: Research, Florida Hospital liaison, committee service, mentoring of graduate and undergraduate students, taught courses in Principals of Modern NMR Spectroscopy, Special Topics in Drug Development, Advanced Biochemistry Laboratory

EARLIER POSITIONS: **Associate Research Scientist** (1991-1993), Yale University School of Medicine, and **Research Associate** (1990-1991), Memorial Sloan-Kettering Cancer Center

---

OTHER EXPERIENCE

GLYGENIX, INC., Cheshire, CT   2005 – 2007
Member, Board of Directors. Glygenix, Inc. was established to benefit children born with Glycogen Storage Disease, Type 1 (GSD1.) Its goal is to help find a cure for this disease by raising monies for GSD1-related research.

THE EPISCOPAL CHURCH AT YALE, New Haven, CT   2000 - 2003
Member, Board of Governors. The Episcopal Church at Yale (ECY) is a full time ministry of the Episcopal Church to students, staff and faculty at Yale.The ECY is governed by a Board of Governors of the Episcopal Church at Yale Corporation which is the legal entity of the Corporation in matters of contracts and other transactions with other institutions such as Yale University.

PUBLICATIONS

Original Articles.

1. Schweitzer, B. I., and Bacopoulos, N. G. (1983) Reversible decrease in dopaminergic 3H-agonist binding after 6-hydroxydopamine and irreversible decrease after kainic acid. *Life Sciences* 32, 531-541.

2. Merker, M., Rice, J., Schweitzer, B., and Handschumacher, R. E. (1983) Cyclosporine binding component in BW5187 lymphoblasts and normal lymphoid tissue. *Transplantation Proceedings* 15, 2265-2270.

3. DeReimer, S. A., Schweitzer, B., and Kaczmarek, L. K. (1985) Inhibitors of calcium-dependent enzymes prevent the onset of afterdischarge in the peptidergic bag cell neurons of Aplysia. *Brain Research* 340, 175-180.

4. Srimatkandada, S., Schweitzer, B. I., Moroson, B. A., Dube, S., and Bertino, J. R. (1989) Amplification of a polymorphic dihydrofolate reductase gene expressing an enzyme with decreased binding to methotrexate in a human colon carcinoma cell line, HCT8R4, resistant to this drug. *Journal of Biological Chemistry* 264, 3524-3528.

5. Schweitzer, B. I., Srimatkandada, S., Gritsman, H., Sheridan, R., Venkataraghavan, R., and Bertino, J. R. (1989) Probing the role of two hydrophobic residues in the active site of the human dihydrofolate reductase by site-directed mutagenesis. *Journal of Biological Chemistry* 264, 20786-20795.

6. Dicker, A. P., Volkenandt, M., Schweitzer, B. I., Banerjee, D., and Bertino, J. R. (1990) Identification and characterization of a mutation in the dihydrofolate reductase gene from the methotrexate resistant Chinese hamster ovary cell line, Pro-3 MTXRIII. *Journal of Biological Chemistry* 265, 8317-8321.

7. Li, W. W., Lin, J. T., Schweitzer, B. I., and Bertino, J. R. (1991) Mechanisms of sensitivity and natural resistance to antifolates in a methylcholanthrene-induced rat sarcoma. *Molecular Pharmacology* 40, 854-858.

8. Li, W. W., Lin, J. T., Chang, Y. M., Schweitzer, B. I., and Bertino, J. R. (1991) Prediction of antifolate efficacy in a rat sarcoma model. *International Journal of Cancer* 49, 234-238.

9. Lin, J. T., Tong, W. P., Trippett, T. M., Niedzwiecki, D., Tao, Y., Tan, C., Steinherz, P., Schweitzer, B. I., and Bertino, J. R. (1991) Basis for natural resistance to methotrexate in human acute non-lymphocytic leukemia. *Leukemia Research* 15, 1191-1196.

10. Sapse, A.-M., Schweitzer, B. I., Dicker, A. P., Bertino, J. R., and Frecer, V. (1992) Ab initio studies of aromatic-aromatic and aromatic-polar interactions in the binding of substrate and inhibitor to dihydrofolate reductase. *International Journal of Peptide and Protein Research* 339, 18-23.

11. Trippett, T., Schlemmer, S., Elisseyeff, Y., Wachter, M., Steinherz, P., Berman, E., Rosowsky, A., Schweitzer, B., and Bertino, J. R. (1992) Evidence for defective transport as a mechanism of acquired resistance to MTX in patients with acute lymphocytic leukemia. *Blood* 80, 1158-1162.

12. Li, W. W., Lin, J. T., Schweitzer, B. I., Tong, W. P., Niedzwiecki, D., Brennan, M. F., and Bertino, J. R. (1992) Intrinsic resistance to methotrexate in human soft tissue sarcoma cell lines. *Cancer Research* 52, 1-6.

13. Li, M.-X., Hantzopoulos, P. A., Banerjee, D., Zhao, S. C., Schweitzer, B. I., Gilboa, E., and Bertino, J. R. (1992) Comparison of the expression of a mutant dihydrofolate reductase under control of different internal promoters in retroviral vectors. *Human Gene Therapy* 3, 381-390.

14. Volkenandt, M., Dicker, A. P., Banerjee, D., Fanin, R., Schweitzer, B. I., Horikoshi, T., Danenberg, K, Danenberg P., and Bertino, J. R. (1992) Quantitation of gene copy number and mRNA using the polymerase chain reaction. *Proceedings of the Society for Experimental Biology and Medicine* 200, 1-6.

15. Fanin, R., Banerjee, D., Volkenandt, M., Waltham, M., Li, W. W., Dicker, A. P., Schweitzer, B. I., and Bertino, J. R. (1993) Mutations leading to antifolate resistance in Chinese hamster ovary cells after exposure to the alkylating agent ethylmethanesulfonate. *Molecular Pharmacology* 44, 13-21.

16. Goker, E., Lin, J. T., Trippet, T., Elisseyeff, Y., Tong, W. P., Niedzwiecki, D., Tan, C., Steinherz, P., Schweitzer, B. I., and Bertino, J. R. (1993) Decreased polyglutamylation of methotrexate in acute lymphoblastic leukemia blasts in adults compared to children with this disease. *Leukemia* 7, 1000-1004.

17. Li, W. W., Waltham, M., Tong, W., Schweitzer, B. I., and Bertino, J. R. (1993) Increased activity of gamma-glutamyl hydrolase in human sarcoma cell lines: a novel mechanism of intrinsic resistance to methotrexate (MTX). *Advances in Experimental Medicine and Biology* 338, 635-638.

18. Kellogg, G. W., and Schweitzer, B. I. (1993) Two- and three-dimensional 31P-driven NMR procedures for complete assignment of backbone resonances in oligodeoxyribonucleotides. *Journal of Biomolecular NMR* 3, 577-595.

19. Dicker, A. P., Waltham, M., Volkenandt, M., Schweitzer, B. I., Otter, G. M., Schmid, F. A., Sirotnak, F. M., and

Bertino, J. R. (1993) Methotrexate resistance in an in vivo mouse tumor resulting from a novel mutation in the dihydrofolate reductase gene. *Proceedings of the National Academy of Sciences* (USA) 90, 11797-11801.

20. Ercikan, E., Waltham, M., Dicker, A., Schweitzer, B., and Bertino, J. R. (1993) Effect of codon 22 mutations on substrate and inhibitor binding for human dihydrofolate reductase. *Advances in Experimental Medicine and Biology* 338, 515-519.

21. Banerjee, D., Schweitzer, B. I., Volkenandt, M., Li, M.-X., Waltham, M., Mineishi, S., Zhao, S.-C., and Bertino, J. R. (1994) Transfection with a cDNA encoding a Ser31 or Ser34 mutant human dihydrofolate reductase into Chinese hamster ovary and mouse marrow progenitor cells confers methotrexate resistance. *Gene* 139, 269-274.

22. Li, M.-X., Banerjee, D., Zhao, S.-C., Schweitzer, B. I., Mineishi, S., Gilboa, E., and Bertino, J. R. (1994) Development of a retroviral construct containing a human mutated dihydrofolate reductase cDNA for hematopoietic stem cell transduction. *Blood* 83, 3403-3408.

23. Zhao, S.-C., Li, M.-X., Banerjee, D., Schweitzer, B. I., Mineishi, S., Gilboa, E., and Bertino, J. R. (1994) Long-term protection of recipient mice from lethal doses of methotrexate by marrow infected with a double-copy vector retrovirus containing a mutant dihydrofolate reductase. *Cancer Gene Therapy* 1, 33-39.

24. Schweitzer, B. I., Mikita, T., Kellogg, G. W., Gardner, K. H., and Beardsley, G. P. (1994) Solution structure of a DNA dodecamer containing the antineoplastic agent cytosine arabinoside: Determination by two- and three-dimensional NMR, restrained molecular dynamics, and complete relaxation matrix analysis. *Biochemistry* 33, 11460-11475.

25. Schweitzer, B. I., Gardner, K. H., and Tucker-Kellogg, G. (1995) HeteroTOCSY-based experiments for measuring heteronuclear relaxation in nucleic acids and proteins. *Journal of Biomolecular NMR* 6, 180-188.

26. Marshalko, S. J., Schweitzer, B. I., and Beardsley, G. P. (1995) Chiral chemical synthesis of DNA containing (S)-9-(1,3-dihydroxy-2-propoxymethyl) guanine (DHPG) and effects on thermal stability, duplex structure, and thermodynamics of duplex formation. *Biochemistry* 34, 9235-9248.

27. Callihan, D., West, J., Schweitzer, B. I., and Logan, T. M. (1996) Simple, distortion-free homonuclear spectra of peptides and nucleic acids in water using excitation sculpting. *Journal of Magnetic Resonance*, Series B, 112, 82-86.

28. Erickan, A. E., Waltham, M., Dicker, A. P., Schweitzer, B. I., Gritsman, H., Banerjee, D,. and Bertino, J. R. (1996) Variants of human dihydrofolate reductase with substitutions at leucine-22: effect on catalytic and inhibitor binding properties. *Molecular Pharmacology* 49, 430-437.

29. Foti, M., Marshalko, S. J., Beardsley, G. P., and Schweitzer, B. I. (1997) Solution structure of a DNA decamer containing the antiviral agent ganciclovir: Determination by two-dimensional homonuclear and heteronuclear NMR, restrained molecular dynamics, and complete relaxation matrix analysis. *Biochemistry* 36, 5336-5345.

30. Kumar, S., Reed, M., Gamper, H., Gorn, V., Lukhtanov, E., Kutyavin, I., Meyer, R., and Schweitzer, B. I. (1997) Solution structure of a DNA decamer conjugated on the 5' end with CDPI3. *Nucleic Acids Research* 26, 831-838.

31. Shao, W., Jerva, L. F., West, J., Lolis, E., and Schweitzer, B. I. (1998) Solution structure of murine macrophage inflammatory protein-2 (MIP-2). *Biochemistry* 37, 8303-8313.

32. Shao, W., Fernandez, E., Wilken, J., Thompson, D. A., Siani, M. A., West, J., Lolis, E., and Schweitzer, B. I. (1998) Accessibility of selenomethionine proteins by total chemical synthesis: structural studies of human herpes virus-8 MIP-II. *FEBS Letters* 441, 77-82.

33. Schweitzer, B. I., Foti, M., Keertikar, K., Kumar, S., Gardner, K. H., and Tucker-Kellogg, G. (1999) The use of ^{31}P relaxation experiments to probe the effects of nucleoside analogs on DNA dynamics. *Phosphorus, Sulfur, and Silicon* 12, 36-41.

34. Foti, M., Omichinski, J. G., West, J., Stahl, S., and Schweitzer, B. I. (1999) Effects of nucleoside analog incorporation on DNA interactions with the DNA binding domain of the GATA-1 erythroid transcription factor. *FEBS Letters* 444, 47-53.

35. Schweitzer, B., Wiltshire, S., Lambert, J., O'Malley, S., Kukanskis, K., Zhu, Z., Ward, D., and Kingsmore, S. F. (2000) Immunoassays with rolling circle DNA amplification: a versatile platform for ultrasensitive antigen detection. *Proceedings of the National Academy of Sciences (USA)* 97, 10113-10119.

36. Wiltshire, S., O'Malley, S., Lambert, J., Kukanskis, K., Edgar, D., Kingsmore, S. F., and Schweitzer, B.I. (2000) Detection of multiple allergen-specific IgE on microarrays by immunoassay with rolling circle amplification. *Clinical Chemistry* 46, 1990-1993.

37. Shao, W., Fernandez, E., Wilken, J., Thompson, D. A., Schweitzer, B. I., and Lolis, E. (2001) CCR2 and CCR5 receptor-binding properties of herpesvirus-8 vMIP-II based on sequence analysis and its solution structure. *European Journal of Biochemistry* 268, 2948-2959.

38. Gusev, Y, Sparkowski, J., Ferguson, H., Montano, J., Visconti, R., Bogdan, N., Schweitzer, B., Wiltshire, S., Kingsmore, S., Maltzman, W., and Wheeler, V. (2001) Rolling circle amplification – A new approach to increase sensitivity for immunohistochemistry and flow cytometry. *Journal of American Pathology* 159, 63-69.

39. Mullenix, M. C., Wiltshire, S., Shao, W., Kitos, G., and Schweitzer, B. (2001) Allergen-specific IgE detection on microarrays using rolling circle amplification: correlation with in vitro assays for serum IgE. *Clinical Chemistry* 47, 1926-1929.

40. Nallur, G., Luo, C., Fang, L., Dave, V., Lambert, J., Lasken, R., Kingsmore, S., and Schweitzer, B. (2001) Signal amplification by rolling circle amplification on DNA microarrays. *Nucleic Acid Research* 29, e118.

41. Schweitzer, B., Roberts, S., Grimwade, B., Shao, W., Wang, M., Fu, Q., Shu, Q., Laroche, I., Christiansen, J., Velleca, M. and Kingsmore, S. F. (2002) Multiplexed protein profiling on microarrays by rolling circle amplification: Application to dendritic cell cytokine secretion. *Nature Biotechnology* 20, 259-365.

42. Nallur, G., Marrero, R., Luo, C., Krishna, R., Bechtel, P., Shao, W., Ray, M., Wiltshire, S, Fang, L, Huang, H., Liu, C.-G., Sun, L., Sawyer, J., Kingsmore, S., Schweitzer, B., and Xia, J. (2003) Protein and nucleic acid detection by rolling circle amplification on gel-based microarrays. *Biomedical Microdevices* 5, 115-123

43. Michaud, G, Salcius, M., Zhou, F., Bangham, R., Bonin, J., Guo, H., Snyder, M., Predki, P., and Schweitzer, B. (2003) Analyzing antibody specificity with whole proteome microarrays. *Nature Biotechnology* 21, 1509-1512.

44. De Yang, Q., Rosenberg, H., Rybak, S., Newton, D., Wang, Z., Fu, Q., Tchernev, V., Wang, M., Schweitzer, B., Kingsmore, S., Patel, D., Oppenheim, J., and Howard. O (2004) Human ribonuclease A superfamily members, eosinophil-derived neurotoxin (EDN) and pancreatic ribonuclease (hPR), induce dendritic cell maturation and activation. *Journal of Immunology* 173, 6134-6142.

45. Predki, P.F., Mattoon, D., Bangham, R., Schweitzer, B., and Michaud, G. Protein microarrays: A new tool for profiling antibody cross-reactivity. *Human Antibodies*. 2005, 14:7-15.

46. Ptacek, J., Devgan, G., Michaud, G., Zhu, H., Zhu, X., Fasolo, J., Guo, H., Jona, G., Breitkreutz, A., Sopko, R., McCartney, R., Schmidt, M., Rachidi, N., Lee, S.-J., Mah, A., Meng, L., Stark, M., Stern, D., De Virgilio, C., Tyers, M., Andrews, B., Gerstein, M., Schweitzer, B., Predki, P., and Snyder., M. (2005) Global analysis of protein phosphorylation in yeast. *Nature* 438, 679-84.

47. Boyle, S.N., Michaud, G.A., Schweitzer B., Predki, P.F., and Koleske, A.J. (2007) A critical role for cortactin phosphorylation by Abl-family kinases in PDGF-induced dorsal-wave formation. *Current Biology* 17, 445-51.

48. Meng, L., Michaud, G.A., Merkel, J.S., Zhou. F., Huang. J., Mattoon, D.R., and Schweitzer, B. (2008) Protein kinase substrate identification on functional protein arrays. *BMC Biotechnology.* 8, 22-30.

49. Schmid, K., Keasey, S., Pittman, P., Emerson, G.L., Meegan, J., Tikhonov, A.P., Chen, G., B. Schweitzer, B., and Ulrich, R.G. (2008) Analysis of the human immune response to vaccinia by use of a novel protein microarray suggests that antibodies recognize less than 10% of the total viral proteome. *Proteomics Clinical Applications* 2, 1528-1538.

50. Sboner, A., Karpikov, A., Chen, G., Smith, M., Mattoon, D., Freeman-Cook, L., Schweitzer, B., Gerstein, M.B. (2009) Robust linear model normalization to reduce technical variability in functional protein microarrays. *Journal of Proteome Research* 8, 5451-5464.

51. Keasey, S.L., Schmid, K.E., Lee, M.S, Meegan, J., Tomas, P., Minto, M., Tikhonov, A.P., Schweitzer, B., Ulrich, R.G. (2009) Extensive antibody cross-reactivity among infectious gram-negative bacteria revealed by proteome microarray analysis. *Molecular and Cellular Proteomics* 8, 924-935.

Reviews.

1. Bertino, J. R., Romanini, A., Dicker, A., Volkenandt, M., Lin, J. T., and Schweitzer, B. I. (1990) Resistance to methotrexate in experimental models and patients. In: Chemistry and Biology of Pteridines 1989, H. C. Curtius, S. Gisal, and N. Blau, Eds. New York: W. deGruyter, 1089-1100.

2. Schweitzer, B. I., Dicker, A. P., and Bertino, J. R., (1990) Dihydrofolate reductase as a therapeutic target. *FASEB Journal* 4, 2441-2452.

3. Bertino, J. R., Romanini, A., Dicker, A., Volkenandt, M., Lin, J. T., and Schweitzer, B. I. (1990) Resistance to MTX: experimental and clinical. In: Drug Resistance: Mechanisms and Reversal. E. Mihich, Ed. New York: John Libbey, 33-46.

4. Bertino, J. R., Li, W. W., Lin, J., Trippett, T., Goker, E., Schweitzer, B., and Banerjee, D. (1992) Enzymes of the

thymidylate cycle as targets for chemotherapeutic agents: mechanisms of resistance. *Mount Sinai Journal of Medicine* 59, 391-395.

5. Banerjee, D., Zhao, S. C., Li, M.-X., Schweitzer, B. I., Mineishi, S., and Bertino, J. R. (1994) Gene therapy utilizing drug resistance genes: a review. Stem Cells 12, 378-385.

6. Schweitzer, B., and Kingsmore, S. (2001) Combining nucleic acid amplification and detection. *Current Opinion in Biotechnology* 12, 21-27.

7. Schweitzer, B., and Kingsmore, S. (2002) Measuring proteins on microarrays. *Current Opinion in Biotechnology* 13, 14-19.

8. Schweitzer, B., Predki, P. and Snyder, M. (2003) Microarrays to characterize protein interactions on a whole-proteome scale. *Proteomics* 3, 2190-2199.

9. Zhou FX, Bonin J, and Predki PF (2004) Development of functional protein microarrays for drug discovery: progress and challenges. *Comb Chem High Throughput Screen* 7:539-46.

10. Merkel JS, Michaud GA, Salcius M, Schweitzer B, and Predki PF. (2005) Functional protein microarrays: just how functional are they? *Curr Opin Biotechnol.* 16:447-52.

11. Bangham R, Michaud GA, Schweitzer B, Predki PF. (2005) Protein microarray-based screening of antibody specificity. *Methods Mol Med.* 114:173-82.

12. Mattoon D, Michaud G, Merkel J, Schweitzer B. (2005) Biomarker discovery using protein microarray technology platforms: antibody-antigen complex profiling. *Expert Reviews in Proteomics* 2:879-89.

13. Michaud, G., Samuels, M., and Schweitzer B. (2006) Functional protein arrays to facilitate drug discovery and development. *IDrugs.* 9:266-72.

14. Michaud, G.A., Salcius, M., Zhou, F., Papov, V.V., Merkel, J., Murtha, M., Predki, P., and Schweitzer, B. (2006) Applications of protein arrays for small molecule drug discovery and characterization. *Biotechnol Genet Eng Rev.* 22:197-211.

15. Salcius, M., Michaud, G.A., Schweitzer, B., and Predki, P.F. (2007) Identification of small molecule targets on functional protein microarrays. *Methods Mol Biol.* 382:239-48.

16. Mattoon, D.R., and Schweitzer, B. (2009) Antibody specificity profiling on functional protein microarrays. *Methods Bol. Biol.* 524: 213-23

EXHIBIT B

Protein-Protein Interaction Profiling on Invitrogen ProtoArray™ High-Density Protein Microarrays

A powerful means of determining the function of a protein is to map its interactions with other proteins. A variety of approaches are available to study protein-protein interactions, including mass spectroscopy, and yeast two-hybrid methods (1). Yet these technologies have several drawbacks: they are time-consuming, require expensive and specialized equipment as well as considerable expertise to run the equipment, and utilize large amounts of sample. Several large-scale efforts to map protein-protein interactions using mass spectroscopy or yeast two-hybrid have been performed recently (2, 3). Interestingly, a comparison of the results of these studies shows little overlap between the interactions observed in each, suggesting that the accuracy or the coverage of the methods may be lacking (4).

Protein microarrays have introduced a new approach to identify and characterize protein interactions, providing the ability to rapidly identify new interactions between thousands of proteins in a single experiment (5). Since the location and identity of each protein on the array is known, interaction maps can be developed rapidly from iterative probings of protein arrays. Because a protein microarray experiment is performed within a day, and interactions are assessed in the context of thousands of other proteins, interaction profiling on microarrays can greatly accelerate the rate at which novel protein interactions are discovered. Additionally, the *in vitro* nature of protein microarray experiments permits control over probing conditions that affect protein interactions such as protein concentration, post-translational modifications, and presence of cofactors, which may not be possible with other methods such as yeast two-hybrid screening.

MacBeath and Schreiber were among the first to demonstrate the potential of protein microarrays in protein-protein interaction, biochemical, and drug binding studies. In this study, pairs of proteins that were known to interact with each other—protein G and the immunoglobulin (IgG), p50 and IκKα, and the FKBP12 binding domain of FKBP with the human immunophilin FKBP12—were shown to interact on protein microarrays (6). Although this study represented a critical milestone in the development of functional protein arrays, only a few proteins were analyzed and novel activities were not identified. Since this report, a series of publications have demonstrated that proteins can retain their expected interactions while immobilized on microarray surfaces. Espejo *et al.* demonstrated that protein interaction domains, such as Src homology (SH2), 14.3.3, forkhead-associated (FHA), PDZ, pleckstrin homology (PH), and FF domains arrayed onto nitrocellulose-coated microarrays retain function and specificity, interacting with their corresponding ligands (7). Newman and Keating have used microarrays to characterize binary coiled-coil interactions from human basic-region leucine zipper transcription factors (8). More recently, Ramachandran *et al.* used protein microarrays to map pairwise interactions among several human DNA replication initiation proteins (9). Finally, in what may be the most striking example of the power of protein microarrays, Michael Snyder and colleagues at Yale University reported the fabrication of an array containing the majority of proteins from the yeast proteome and the use of this array to identify a new binding motif for calmodulin (10).

INVITROGEN PROTOARRAY™ PRODUCTS

Invitrogen has recently introduced the ProtoArray™ Microarray Technology for studying molecular interactions on protein arrays. The ProtoArray™ products include the ProtoArray™ Yeast Proteome Microarray nc v1.0, which contains 4088 open reading frames (ORFs) from *Saccharomyces cerevisiae*, and the ProtoArray™ Human Protein Microarray nc v1.0, which consists of nearly 1,900 human proteins. All proteins are expressed as N-terminal glutathione-S-transferase (GST) fusion proteins, purified, and spotted in duplicate on nitrocellulose-coated 1 inch x 3 inch glass slides. Using ProtoArray™ Microarrays allows screening of target proteins of interest for interaction with thousands of proteins in as little as four hours. Detection on the arrays is sensitive—as little as 1 pg protein on the array can be detected with submicrogram quantities of probe protein—and reproducible.

To detect protein-protein interactions on ProtoArray™ Microarrays, the protein probe must contain a label or tag to visualize the interaction of the probe with array proteins. The extremely high affinity of the biotin-streptavidin interaction makes biotin-protein conjugation a preferred method for protein labeling. Invitrogen offers the ProtoArray™ PPI Complete Kit for biotinylated proteins, which contains a module for efficiently biotinylating small amounts of a protein as well as qualified reagents for blocking, washing, and detecting biotinylated protein probes with streptavidin conjugated to a fluorescent dye, Alexa Fluor 647.

Another preferred method of detecting protein interactions on ProtoArray™ Microarrays is to use protein probes with an epitope tag and a labeled antibody against the tag. An example of such a tag is the V5 epitope, a 14 amino acid (GKPIPNPLLGLDST) epitope derived from the P and V proteins of the paramyxovirus SV5. Invitrogen offers several Gateway expression vectors that allow the fusion of the V5-tag to a protein of interest. The ProtoArray™ PPI Complete Kit for epitope-tagged proteins from Invitrogen provides reagents for blocking, washing, and detecting a V5-tagged protein using an Anti-V5-

Alexa Fluor 647 Antibody developed specifically for this application.

This Application Note demonstrates the utility of Yeast and Human ProtoArray™ Protein Microarrays for detecting protein-protein interactions using the biotinylated or epitope-tagged protein probes.

MATERIALS AND METHODS

Yeast Proteome collection: The yeast proteome collection was derived from the yeast clone collection of yeast ORFs generated by the Snyder laboratory as described by Zhu *et al.* (10). Each *S. cerevisiae* open reading frame (ORF) was expressed as an N-terminal GST-6xHis fusion protein in a yeast expression vector. The identity of each clone was verified using 5′-end sequencing and the expression of GST-tagged fusion protein by each clone was confirmed with Western immunodetection using an anti-GST antibody. After verifying that each clone expresses a protein of the expected molecular weight, the proteins (from 4,088 clones) were expressed and purified using high-throughput procedures (10).

Human protein collection: The majority of the human protein collection is derived from the human Ultimate™ ORF Clone Collection available from Invitrogen (see http://orf.invitrogen.com for more information). The human proteins were expressed in the Bac-to-Bac Baculovirus Expression System (Invitrogen Cat. no. 10359-016, for more information on the Bac-to-Bac Baculovirus Expression System, visit www.invitrogen.com). Each Ultimate™ ORF Clone (entry clone) consists of a human ORF cloned into a Gateway entry vector. Each entry clone was subjected to an LR reaction with the Gateway destination vector, pDEST™20 to generate an expression clone. The LR reaction mix obtained after performing the LR reaction was transformed into competent DH10Bac™ *E. coli* to generate a recombinant bacmid. The high molecular weight recombinant bacmid DNA was isolated and transfected into Sf9 insect cells to generate a recombinant baculovirus that was used for preliminary expression

experiments. After the baculoviral stock was amplified and titered, the high-titer stock was used to infect Sf9 insect cells for expression of the recombinant protein of interest in 96 deep-well plates. Following a 3-day growth, the insect cells were harvested for purification. All steps of the purification process including cell lysis, binding to affinity resins, washing, and elution, were carried out at 4°C. Insect cells are lysed under non-denaturing conditions and lysates were loaded directly into 96-well plates containing glutathione resin. After washing, purified proteins were eluted under conditions designed to obtain native proteins. After purification, samples of the purified proteins were run in SDS-PAGE gels and immunodetected by Western blot. The gel images were processed to generate a table of all the protein molecular weights detected for each sample.

ProtoArray™ manufacturing: The protein purification process described above produces thousands of purified proteins ready to be printed on arrays. A contact-type printer equipped with 48 matched quill-type pins is used to deposit each of these proteins along with a set of control elements in duplicate spots on 1" x 3" glass slides. The printing of these arrays is performed in a cold room under dust-free conditions to preserve the integrity of protein samples and printed microarrays. Before releasing the protein microarrays for use, each lot of arrays is subjected to a rigorous quality control procedures, including visual inspection of all the printed arrays to check for scratches, fibers, smearing, etc. To control for the quality of the printing process, several microarrays are probed with an anti-GST antibody. Since each protein contains a GST fusion tag, this procedure measures the variability in spot morphology, the number of missing spots, the presence of control spots, and the amount of protein deposited in each spot.

Cloning, Expression, and Purification of Proteins (6xHis-V5-BioEase™-EK-protein fusions): Ultimate™ ORF clones were obtained as entry clones and L x R cloned into pET105 for expression in *E. coli*. For each ORF, plasmid DNA was transformed into BL21 Star™ (DE3) *E. coli* cells, which were plated on LB/Amp and grown overnight at 37°C. Several colo-

3

nies from each of the 12 constructs were picked from LB/Amp plates and transferred into 50 ml of LB Amp. Cultures were grown from 5 to 7 hours at 37°C until an OD_{600} of 0.5 to 0.6 was reached. Next, 50 µl of 0.1 M IPTG was added to give a final concentration of 100 µM, and these cultures were incubated overnight at 20°C. Cell lysates were prepared using the protocol described in the ProBond™ Purification Resin manual. Pellets were resuspended with 8 ml Native Binding Buffer; 8 mg lysozyme was added and lysed for 30 minutes on ice. Cells were then sonicated on ice with six 10-second bursts and then centrifuged at 3,500 rpm for 20 minutes. Lysate (8 ml) was loaded onto a column with 2 ml washed ProBond™ resin and incubated for 1–2 hours at 4°C. The column was washed with Native Wash Buffer followed by an elution with 10 ml Elution Buffer. The pooled fractions were dialyzed twice against 2 L PBS. All samples were concentrated on Millipore spin membrane cartridges (10,000 MW cut-off) to a final volume of 250–350 µl, and were brought to 5% glycerol by the addition of an appropriate amount of 100% glycerol. Samples were then quick-frozen in liquid nitrogen and stored at −80°C.

In vitro biotinylation of proteins: Human calmodulin (Upstate) was biotinylated using the protocol outlined in the ProtoArray™ Mini-Biotinylation Kit (Invitrogen). Briefly, protein was biotinylated at room temperature for 1 hour and the sample was applied to a gel filtration column to remove unincorporated biotin. Protein concentration and the extent of labeling was also assessed.

Alexa Fluor® 647-streptavidin based detection: The protein-protein interaction assay was performed using the protocol outlined in the ProtoArray™ PPI Complete Kit for biotinylated proteins (Invitrogen). Arrays were blocked with 1% BSA/PBST at 4°C for 1 hour. Proteins were diluted in probe buffer (1X PBS, 5 mM MgCl₂, 0.5 mM DTT, 5% glycerol, 0.05% Triton X-100, 1% BSA) to 5 or 50 ng/µl and added to arrays under a cover slip, Hybrislip (included in the kit). Proteins were incubated at 4°C for 90 minutes in a 50 ml conical tube and then transferred to an incubation/hybridization chamber (included with the kit). Arrays were washed three times with probe

buffer. Subsequently, a solution of Alexa Fluor® 647-streptavidin (Invitrogen, 0.75 µg/ml) in probe buffer was added and incubated at 4°C for 30 minutes. Arrays were washed three times and dried.

Anti-V5-Alexa Fluor® 647 based detection: The protein-protein interaction assay was performed using the protocol outlined in the ProtoArray™ PPI Complete Kit for epitope-tagged proteins (Invitrogen). Arrays were blocked with 1% BSA/PBST at 4°C for 1 hour. Proteins were diluted in probe buffer to 5 or 50 ng/µl and added to arrays under a Hybrislip cover slip. Proteins were incubated at 4°C for 90 minutes in a 50 ml conical tube and then transferred to an incubation/hybridization chamber (included in the kit). Arrays were washed three times with probe buffer. Subsequently, a solution of anti-V5-Alexa Fluor® 647 conjugated antibody (Invitrogen, 0.25 µg/ml) was added and incubated at 4°C for 30 minutes. Arrays were washed three times and dried.

Data acquisition/analysis: The microarray was scanned with a GenePix® 4000B Fluorescent Scanner (Molecular Devices). Data was acquired with GenePix® Pro software (Molecular Devices) and processed using ProtoArray™ Prospector (a software tool developed by Invitrogen that automatically performs data analysis, see www.invitrogen.com/protoarray for details)

or Microsoft Excel and Microsoft Access. Statistically significant signals on each protein array were identified. The significant signals are greater than or equal to a value that is determined by calculating the median plus three standard deviations (using signal minus background values for all non-control proteins) for all non-control proteins on the array. Interactors were defined as proteins having positive significance calls not observed on the appropriate negative control.

RESULTS

Probing ProtoArray™ Yeast Proteome Microarrays with biotinylated yeast proteins: Four yeast proteins were biotinylated *in vitro* using the Invitrogen ProtoArray™ Mini Biotinylation Kit. As shown in Figure 1, all four proteins showed expected interactions when used to probe the ProtoArray™ Yeast Proteome Microarray and detected with Alexa Fluor® 647-Streptavidin Conjugate. Each of the identified interactions is well annotated in the literature using a variety of different approaches (see http://www.yeastgenome.org for further details). Note that the interactions shown in Figure 1 are reciprocal. Biotinylated Ybr109C (calmodulin) interacts with Yfr014C (calmodulin kinase) on the array, and biotinylated Yfr014C interacts with Ybr109C on the array; the same relationship is observed with the GTP binding protein
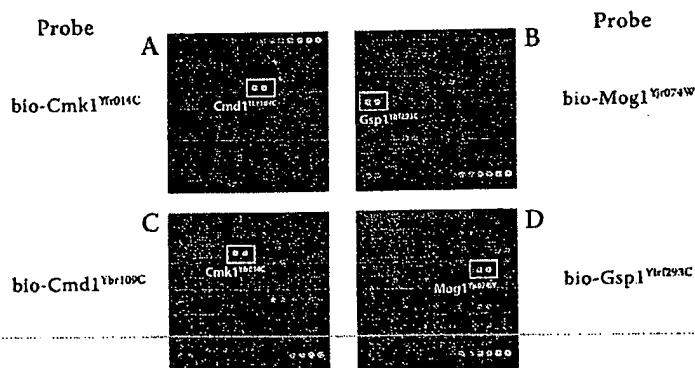
Figure 1—Probing the ProtoArray™ Yeast Proteome Microarray with *in vitro* biotinylated yeast proteins. Subarrays show expected interactions with biotinylated yeast proteins. Proteins were concentrated to 250 µg/ml and biotinylated using the ProtoArray™ Mini-Biotinylation Kit.

Gsp1 (Ylr293C) and the nuclear transport protein Mog1 (Yjr074W). The reciprocal interactions are important for demonstrating the validity of the observed interactions and the functionality of the proteins on the array.

Probing ProtoArray™ Human Protein Microarrays with Biotinylated and Epitope-tagged Human Proteins: To assess the utility of human protein arrays and protein-protein interaction detection technologies optimized at Invitrogen for demonstrating protein-protein interactions, proteins containing both a single biotin and a V5 tag were prepared (see Materials and Methods). Several N-terminal fusions of V5-BioEase™ human proteins were probed against human protein arrays (ProtoArray™ Human Protein Microarray nc v1.0) consisting of approximately 1,900 purified human proteins spotted in duplicate on a nitrocellulose-coated glass slide. After probing the array with calmodulin 2 (CALM2), we observed that CALM2 interacted with several proteins on the array. Most notable are the interactions with calcium/calmodulin-dependent protein kinase IV (CAMK4) and calcium/calmodulin-dependent protein kinase I (CAMK1) (Figure 2). These interactions were observed when streptavidin (data not shown) or anti-V5 based detection was used (Figure 2). We also used the ProtoArray™ Mini-Biotinylation Kit to in vitro biotinylate

recombinant human calmodulin and used this protein to probe the ProtoArray™ Human Protein Microarray nc v1.0. As shown in Figure 3, similar protein interactions with CAMK1 and CAMK4 were observed for in vitro biotinylated calmodulin as with the BioEase™-tagged CALM2, demonstrating that valid protein-protein interaction data can be obtained by using proteins that are biotinylated using in vitro or in vivo methods.

To demonstrate the utility and ease of use of ProtoArray™ Technology for identifying novel protein-protein interactions, a V5-BioEase™ fusion to the protein cyclin-dependent kinase inhibitor 1B (CDK1NB, p27, Kip1) was used to probe a ProtoArray™ Human Protein Microarray. We identified a specific interaction with cyclin-dependent kinase 7 (Cdk7, MO15 homolog, *Xenopus laevis*, cdk-activating kinase) (Figure 4). The same interaction was also observed using streptavidin-based detection (data not shown). Although this interaction has not been reported previously in the literature, an interaction of CDK1NB with Cdk3 has been reported, and it has been proposed that retinoic acid induces cell cycle arrest in tumor cell lines by promoting formation of this complex (11). To validate the interaction, we performed the following reciprocal protein-protein interaction assay: CDK1NB was spotted on a nitrocellulose coated slide, then probed

with GST-tagged Cdk7, and the Cdk7-CDK1NB complex was detected using an anti-GST antibody. Similar probings with 18 other GST-tagged proteins gave signals with the spotted CDK1NB that were on average approximately 10-fold lower than Cdk7 (Figure 5), indicating that the Cdk7-CDK1NB interaction is quite specific.

SUMMARY

ProtoArray™ Protein Microarrays with Alexa Fluor® detection technologies are optimized to quickly identify novel protein-protein interactions. High-quality reagents, protocols and technical support are available. Consult the Invitrogen website for the latest information regarding protein microarrays for protein interaction profiling using ProtoArray™ Technology.
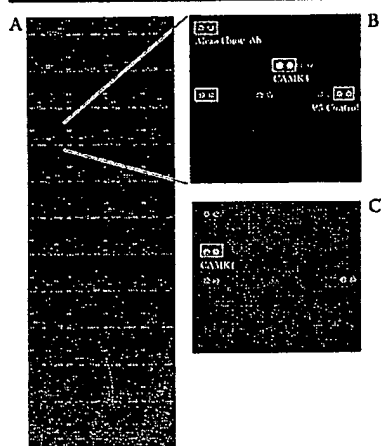
◀ Figure 2—ProtoArray™ Human Protein Microarray nc v1.0 probed with human CALM2. Interactions detected with anti-V5-Alexa Fluor® 647 Dye.

Panel A Whole slide image.

Panel B Interaction of CALM2 with CAMK4. Signals from Alexa Fluor® Antibody and V5 control are shown. Alexa Fluor® labeled antibody is in every subarray and used as a reference marker for aligning the data acquisition grid. The V5 control is a V5 tagged protein printed on the slide. Signal with this protein indicates that assay detection is functioning properly.

▶ Figure 3—ProtoArray™ Human Protein Microarray nc v1.0 probed with *in vitro* biotinylated human calmodulin. Interactions detected with streptavidin Alexa Fluor® 647 Dye.

Panel A Whole slide image.

Interactions of human calmodulin with CAMK4 (*Panel B*) and CAMK1 (*Panel C*). Signals from Alexa Fluor® Antibody and biotinylated antibody gradient are shown. The biotinylated antibody gradient is used as assay detection control. Signal with this protein indicates that assay detection is functioning properly.
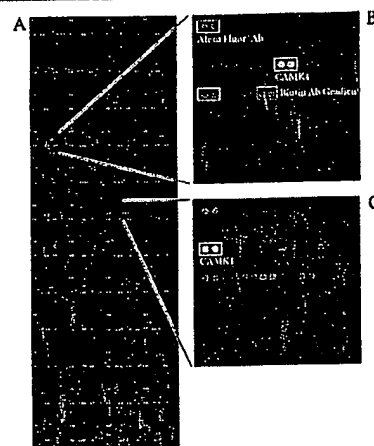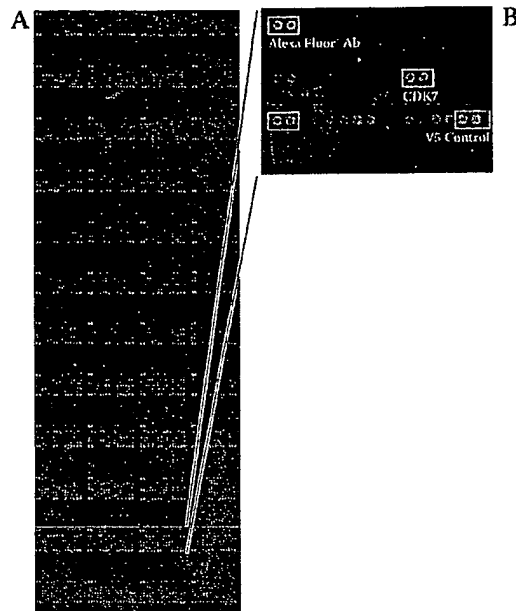
Figure 4—ProtoArray™ Human Protein Microarray nc v1.0 probed with CDKN1B. Interactions detected with anti-V5-Alexa Fluor® 647 Dye. *Panel A* Whole slide image. *Panel B* Interaction of CDKN1B with CDK7. Signals from Alexa Fluor® Antibody and V5 control are shown.
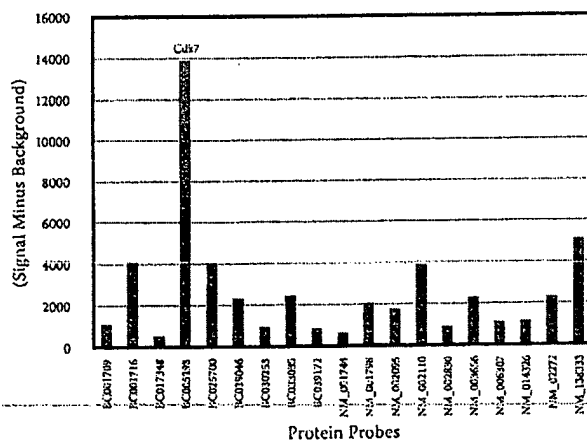
Figure 5—Reciprocal Protein Interaction Assay. Nineteen GST-fusions were expressed in Sf9 cells, purified using glutathione chromatography, and probed against an array containing immobilized CDKN1B. The Y-axis is the signal background value for the CDKN1 spot for each protein probed (X-axis) against the array. The accession numbers (MGC or RefSeq) for the protein probes are listed (X-axis). The MGC accession number for Cdk7 is BC005298. CDKN1B was spotted at an equivalent solution protein concentration of approximately 12 ng/μl. The median probing concentration for the 19 proteins was 11 ng/ul; the mean concentration was 12 ng/μl.

REFERENCES

1. Michaud, G. A. & Snyder, M. Proteomic approaches for the global analysis of proteins. *Biotechniques* 33, 1308–16 (2002).

2. Uetz, P. *et al.* A comprehensive analysis of protein-protein interactions in Saccharomyces cerevisiae. *Nature* 403, 623–7 (2000).

3. Ito, T. *et al.* A comprehensive two-hybrid analysis to explore the yeast protein interactome. *Proc Natl Acad Sci U S A* 98, 4569–74 (2001).

4. von Mering, C. *et al.* Comparative assessment of large-scale data sets of protein-protein interactions. *Nature* 417, 399–403 (2002).

5. Schweitzer, B., Predki, P. & Snyder, M. Microarrays to characterize protein interactions on a whole-proteome scale. *Proteomics* 3, 2190–9 (2003).

6. MacBeath, G. & Schreiber, S. L. Printing proteins as microarrays for high-throughput function determination. *Science* 289, 1760–3 (2000).

7. Espejo, A., Cote, J., Bednarek, A., Richard, S. & Bedford, M. T. A protein-domain microarray identifies novel protein-protein interactions. *Biochem J* 367, 697–702 (2002).

8. Newman, J. R. & Keating, A. E. Comprehensive identification of human bZIP interactions with coiled-coil arrays. *Science* 300, 2097–101 (2003).

9. Ramachandran, N. *et al.* Self-assembling protein microarrays. *Science* 305, 86–90 (2004).

10. Zhu, H. *et al.* Global analysis of protein activities using proteome chips. *Science* 293, 2101–5 (2001).

11. Hsu, S. L., Hsu, J. W., Liu, M. C., Chen, L. Y. & Chang, C. D. Retinoic acid-mediated G1 arrest is associated with induction of p27 (Kip1) and inhibition of cyclin-dependent kinase 3 in human lung squamous carcinoma CH27 cells. *Exp Cell Res* 258, 322–31 (2000).

EXHIBIT C

Development and Validation of Kinase Substrate Screening on ProtoArray™ High-Density Protein Microarrays

B. Schweitzer, G. Michaud, R. Bangham, J. Bonin, M. Salcius, and P. Predki

Abstract

Identifying biologically relevant substrates for protein kinases is a critical step in understanding the function of these clinically important enzymes. Traditional approaches for kinase substrate identification are expensive, slow, and lack sensitivity. For this reason, many kinase activity assays employ generic substrates or peptides that decrease the reliability of these assays for drug development. We describe here the development and validation of a rapid and sensitive microarray-based kinase substrate identification technology, which enables parallel screening of kinases against thousands of potential native protein substrates. This paper describes the validation of this approach and use of the resulting data for pathway mapping.

Introduction

Protein kinases play a central role in the regulation of multiple cellular processes and in diseases; in fact, 244 kinases have been mapped to disease loci (1). It is not surprising, therefore, that a large number of biotech and pharmaceutical companies are seeking to discover and bring to the clinic compounds that demonstrate specific inhibition of kinases involved in disease. Some examples of kinase inhibitors already in clinic include Gleevec® (Novartis), an Abl and c-Kit kinase inhibitor that has been successful in the treatment of chronic myeloid leukemia and gastrointestinal stromal tumors, and Herceptin® (Genentech), an antibody that targets the HER2/neu (erbB2) protein for treatment of breast cancer. The family of human protein kinases consists of more than 500 members of which only a fraction have been characterized to date. Much is still not known about the biological function of many kinases, the protein substrates that are phosphorylated by these kinases, or the roles of these kinases and substrates in disease.

The importance of protein kinases in virtually all processes regulating cell transduction illustrates the potential for kinases and their cellular substrates as targets for therapeutics. Considerable efforts have been made to elucidate kinase biology by identifying the substrate specificity of kinases and using this information for the prediction of new substrates. Some of the approaches used to date include creation of a database from annotated phosphorylation sites, prediction of substrate sequence patterns from available structures of kinase/peptide substrate complexes, and screening of peptide libraries and peptide arrays (2,3). More recent efforts include attempts to map the phosphoproteome using mass spectroscopy-based techniques. While these studies have provided some information about kinase biology, they have been severely limited by their complexity, expense, lack of sensitivity, the use of non-structured peptides, and by poor representation of potential substrates in the screens.

Invitrogen is pioneering the use of arrays of whole or partial proteomes to improve the success rates of drug discovery. This report describes how ProtoArray™ technology rapidly converts gene sequences into arrays of functional proteins that can be used to reveal new disease pathways and define the specificity and selectivity of potential drugs. In addition, this paper discusses how the ProtoArray™ high-density protein microarray technology is an ideal format for identifying biologically relevant substrates for protein kinases in a rapid, cost-effective, and comprehensive fashion.

Validation results

ProtoArray™ technology enables fast, simple, and comprehensive kinase substrate screening.

Each ProtoArray™ microarray contains thousands of *S. cerevisiae* or *H. sapiens* proteins spotted in high density on glass slides. These slides can be probed to identify protein interactions with DNA, proteins, lipids, sugars, small molecules, and enzymes. The first proof-of-principle experiment demonstrating that these arrays can be used to reveal substrates of proteins kinases was carried out on the Yeast ProtoArray™ microarray, which contains over 4000 unique yeast proteins spotted in duplicate. The experimental outline is simple (Figure 1A). A solution comprising a kinase and radioactive ATP was incubated on a Yeast ProtoArray™ microarray, and then the slide was washed and exposed to a phosphoimager (Figure 1B). The experiment identified 41 proteins specifically phosphorylated by the exogenous kinase.

Figure 1. Kinase-substrate assay on the Yeast ProtoArray™ Microarray.
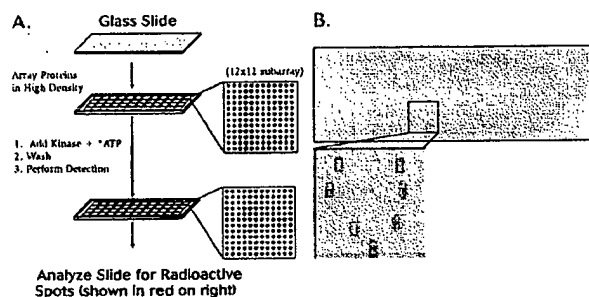
A) Experimental design of substrate screening assay. B) The Yeast ProtoArray™ microarray containing > 4000 different yeast proteins probed with a purified kinase. Inset: positives boxed in green (autophosphorylation) and red (substrates).

Initial work with Human ProtoArray™ microarrays demonstrates kinase substrate discovery value. To test our platform for identification of kinase substrates, we chose the human protein kinase Arg. This kinase, along with its closely related homolog Abl, is known to be involved in the etiology of chronic myeloid leukemia (CML) and is a target for the anti-cancer agent Gleevec®. Human ProtoArray™ microarrays were manufactured with 1500 different quality-controlled recombinant human proteins produced in Invitrogen's proprietary high-throughput insect cell expression and parallel purification systems. A known Abl/Arg substrate, Crk, was printed in regular intervals on the array as a positive control. The Human ProtoArray™ microarray in Figure 2A was incubated with radiolabeled ATP alone; proteins that show a signal on this array are kinases present on the array that autophosphorylate. The array in Figure 2B was incubated with Arg in the presence of radiolabeled ATP. This kinase phosphorylated the control substrate Crk in every subarray; in addition, nine other proteins, that did not give signal with ATP alone, were observed to be phosphorylated in the presence of Arg. We also looked at the effect of adding an Arg/Abl kinase-specific inhibitor and found that the inhibitor specifically decreased phosphorylation of Crk and the nine other microarray identified substrates (Figure 2C), confirming that these proteins were phosphorylated by Arg kinase.

Figure 2. Identification of substrates for Arg kinase.
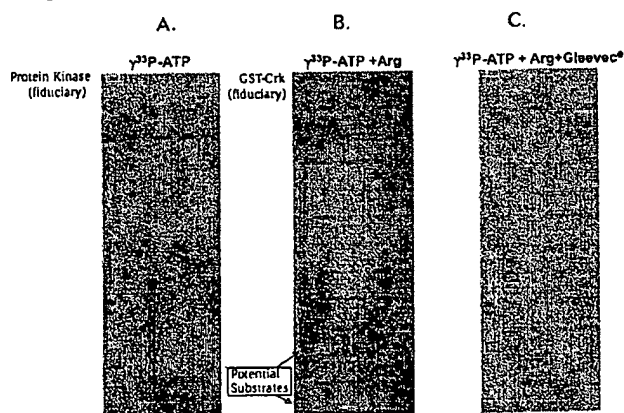
ProtoArray™ microarrays containing 1500 different human proteins were treated with ATP (A), ATP and Arg (B), or ATP, Arg, and Arg-specific inhibitor (C). Nine substrates were identified for Arg (boxed in red).

Verification of specific phosphorylation by a human kinase.

Arg kinase is known to specifically phosphorylate tyrosine residues on certain proteins. To verify that Arg kinase maintains this specificity for tyrosine residues in array-based experiments, Human ProtoArray™ microarrays were treated sequentially with Arg kinase followed by a phosphotyrosine phosphatase. As shown in Figure 3, all proteins phosphorylated by Arg kinase on the array are dephosphorylated by the phosphotyrosine phosphatase, confirming that Arg kinase substrates on the array are appropriately phosphorylated on tyrosine residues. Signals from proteins that autophosphorylate (*i.e.*, that show signal in the absence of exogenous kinase) were not affected by phosphotyrosine phosphatase treatment, indicating that these were kinases that autophosphorylate serine/threonine residues.

Figure 3. Phosphotyrosine phosphatase reduces Arg substrate phosphorylation.
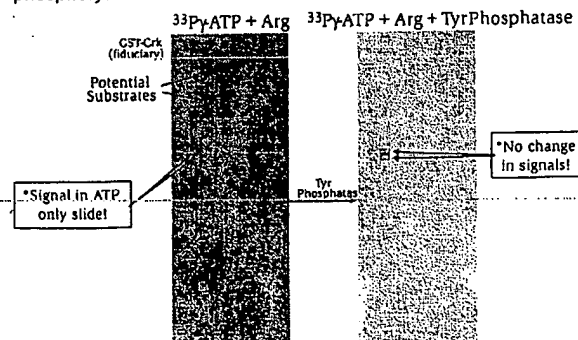
A Human ProtoArray™ microarray containing the eight identified Arg substrates was probed with Arg kinase and then subsequently with a phosphotyrosine phosphatase.

Substrate phosphorylation is kinase-specific. The results with Arg kinase on Human ProtoArray™ microarrays clearly demonstrated that this kinase is highly selective in the protein substrates that it phosphorylates. In order for this application of the ProtoArray™ technology to be useful to a wide range of kinase biologists, the ability to distinguish phosphorylation patterns of different kinases must be established. Consequently, ProtoArray™ microarrays printed with 2500 different human proteins were incubated with ^{33}P-ATP and either Arg or PKC kinase (Figure 4) or with ^{33}P-ATP alone (not shown). As shown in Figure 4, phosphorylation signals specific to each kinase were clearly observed. The majority of signals present in both experiments were due to autophosphorylation by some of the ~400 kinases printed on the array. Analysis of the whole array revealed dozens of proteins that were specific to one of the kinases. We have now characterized the phosphorylation patterns of over a dozen different human kinases and have identified large numbers of unique substrates for each kinase.

Figure 4. Specificity of kinase phosphorylation on Human ProtoArray™ microarrays.
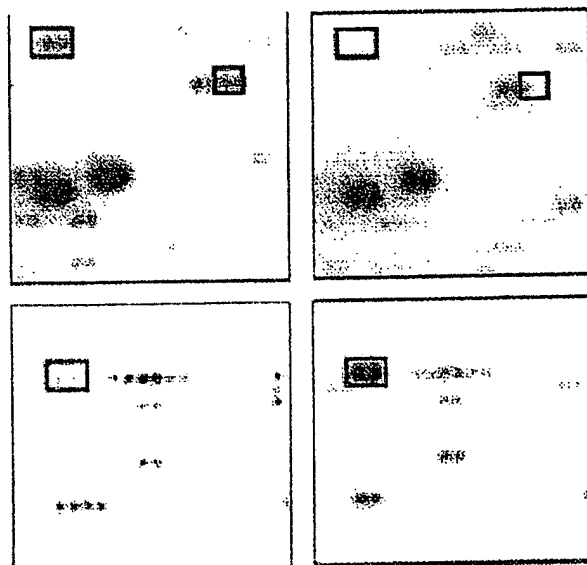
Two Human ProtoArray™ microarrays were incubated with ^{33}P-ATP and either Arg (left column) or PKC kinase (right column). Two representative subarrays are shown: In subarray 1 (top row), two proteins phosphorylated specifically by Arg kinase are boxed in blue; in subarray 2 (bottom row), a protein phosphorylated specifically by PKC is boxed in red.

Validation of substrate identification in an independent assay. Biochemical validation of the array-based substrate screening assay was initially carried out by determining whether proteins phosphorylated by Arg kinase on the array would also be phosphorylated in a different assay format. Figure 5 shows the results of assays in which two of the substrate proteins were incubated in solution with Arg kinase in the presence of radiolabeled ATP. Separation of the reaction mixtures on denaturing gels demonstrated that proteins at the expected molecular weight of the substrate proteins were indeed phosphorylated in solution. These results strongly suggest that these proteins maintain their native conformation on the array, allowing them to be phosphorylated by specific kinases.

Detailed validation studies reveal the highest affinity substrate for a pharmacologically relevant kinase reported to date. Although phosphorylation of proteins by kinases in experiments, such as the one shown in Figure 5, is a prerequisite for identifying substrates for these enzymes, additional lines of evidence are needed to demonstrate physiological relevance. One such line of evidence is data showing that the substrate is phosphorylated at concentrations likely to occur in a cell. One of the eight proteins identified on the ProtoArray™ microarray as a substrate for Arg kinases was selected for more detailed K_M measurements based on the protein's known role in cell division. Analysis of the data from this experiment yields a K_M for the substrate of approximately 50 nM (Figure 6). Not only is this value well within a potential intracellular concentration for a protein, but it is also lower than any K_M value previously reported for Arg kinase.

Figure 5. Arg kinase phosphorylation of substrate in solution.
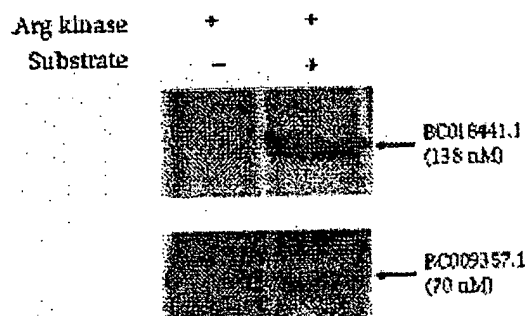
Arg kinase alone or mixed with substrate proteins was incubated at 30°C for 30 minutes and then run on an SDS-PAGE gel and phosphoimaged.

Figure 6. K_M determination for an Arg kinase substrate identified on the ProtoArray™ microarray.
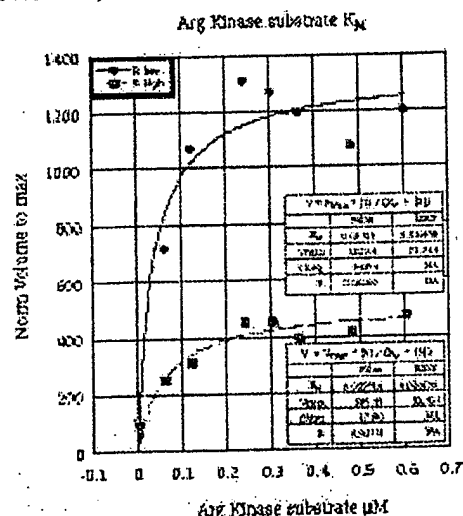
Arg kinase was incubated with different concentrations of the substrate protein and the phosphorylation of the protein was measured in a gel-based assay similar to the one shown in Figure 4.

ProtoArray™ data is used to generate a new kinase pathway. In addition to biochemical validation, it is also desirable to see concordance of ProtoArray™ results with published data. In fact, a search of the literature and publicly available databases revealed that one of the proteins proven to be a substrate for Arg Kinase on a Human ProtoArray™ microarray, Shp1, had indeed been annotated as a substrate for this kinase. Using a protein-protein interaction assay on a Human ProtoArray™ microarray, we also demonstrated for the first time that Arg kinase forms a stable interaction with Shp1 (data not shown). Shp1 is a phosphotyrosine phosphatase localized at the plasma membrane; our data, as well as the published data, are therefore consistent with co-localization and co-regulation of Shp1 phosphatase and Arg kinase (Figure 7). Other published reports indicate that following activation by Src, Arg and Abl kinases translocate into the nucleus, although the functional consequences of this translocation have not been clarified. ProtoArray™ results, however, clearly showed that these kinases phosphorylated several transcription factors that may have roles in cell cycle function. An RNA polymerase was also phosphorylated, providing another line of evidence that these kinases regulate RNA transcription and gene expression. Equally intriguing is the finding that a membrane-associated receptor present on the array was phosphorylated by Arg kinase. Interaction of this receptor with a membrane-associated kinase has been shown by others to result in the activation of two kinases that have been implicated in oncogenesis. This finding represents a new and potentially therapeutically relevant link between the Arg/Abl kinases and cancer.

Figure 7. Pathway mapping with Arg kinase-substrate ProtoArray™ data.
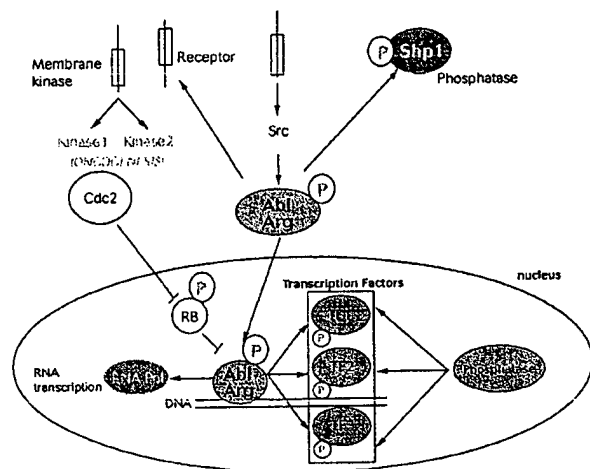
Conclusion

We have combined unprecedented protein content with a simple-to-use microarray assay to generate new knowledge about protein kinases with unequalled efficiency. We have demonstrated specific phosphorylation of both known and novel substrates using Human and Yeast ProtoArray™ high-density protein microarrays and have validated these proteins as substrates using more standard assays. Combining this new type of information with Invitrogen's other capabilities for measuring phosphatase activity, protein-protein interactions, and drug inhibition on microarrays allows scientists to link kinases to intracellular signaling networks and generate new understandings about kinases and their substrates as drug targets with unmatched speed and efficiency.

Implications of ProtoArray™ technology

The discovery of new kinase substrates by Invitrogen and its collaborators using the ProtoArray™ technology platform demonstrates the enormous value of high-content protein arrays. This was clearly illustrated in experiments using Arg kinase: nine substrates were identified using an array printed with 1500 human proteins, but six more were found using a 2500 protein array. Extrapolating to an array containing a representative protein from the approximately 30,000 human genes (the UniProteome) suggests that over 150 substrates would be identified, thereby greatly increasing the informational value of the experiment.

References

1. Manning, G., Whyte, D.B., Martiniez, R., Hunter, T., and Sudarsanam, S (2002). The protein kinase complement of the human genome. *Science 298, 1912-1934.*

2. Kreegipuu, A., Blom, N., Brunak, S. and Jarv, J (1998) The statistical analysis of protein kinase specificity determinants. *FEBS Letters 430, 45-50.*

3. Rychlewski, L., Kschischo, M., Dong, L., Schutkowksi, M., and Reimer, U. (2004) Target specificity analysis of the Abl kinase using peptide microarray data. *J. Mol. Biol. 336, 307-311.*

Invitrogen
life technologies

O-13121-r1 US 0704 xM

EXHIBIT D

Access the human proteome on a microarray scale

ⓘ invitrogen

Ordering information
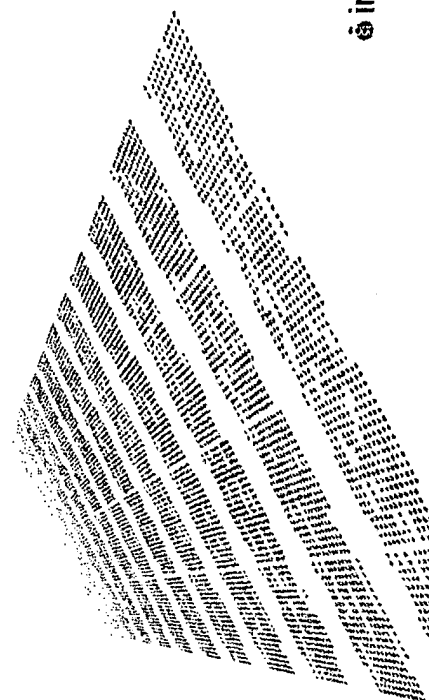
Product	Quantity	Cat. no.
ProtoArray® Human Protein Microarray v4.0 for protein-protein interactions (PPI)	1 array	PAH052401
ProtoArray® Human Protein Microarray PPI Complete Kit for biotinylated protein	1 kit	PAH052401
ProtoArray® Human Protein Microarray PPI Complete Kit for V5 epitope tagged proteins	1 kit	PAH052403
ProtoArray® Human Protein Microarray v4.0 for immune response biomarker profiling (IRBP)	1 array	PAH052402
ProtoArray® Human Protein Microarray Discovery Study Bundle	20 assays	PAH052405
ProtoArray® Human Protein Microarray Discovery Study Bundle	40 assays	PAH052407
ProtoArray® Human Protein Microarray Discovery Study Bundle	60 arrays	PAH052409
Immune Response Biomarker Profiling Application Kit	1 kit	PA016
ProtoArray® Control Protein Microarray v4.0 for protein-protein interactions (PPI)	1 array	PA007
ProtoArray® Human Protein Microarray v4.0 for kinase substrate identification (KSI)	1 array	PAH052405
ProtoArray® Human Protein Microarray KSI Complete Kit for kinase substrate identification	1 kit	PAH052405
ProtoArray® Kinase Substrate Identification Application Kit	1 kit	PA015
ProtoArray® Control Protein Microarray v4.0 for kinase substrate identification (KSI)	1 array	PA002

References

1. Merkel J.S. et al. (2005) Curr Opin Biotech 16:1–6.
2. Robinson, W. H. et al. (2002) Nat Med 8: 295–301.
3. Zhu, H. et al. (2001) Science 293, 2101–2105.
4. Predki, P. (2004) Curr Opin Chem Biol 8: 8–13.
5. Michaud, G. A. et al. (2004) DDT: Targets 3: 238–245.
6. Michaud, G. A. et al. (2003) Nat Biotechnol 21: 1509–1512.

Access the human proteome on a microarray scale

ProtoArray® Human Protein Microarray v4.0

→ Study >8,000 human proteins from multiple gene families

→ Generate biologically relevant results by screening functional proteins

→ Detect femtomole levels of protein from very small sample sizes

→ Identify proteins immediately and access sequence-verified Gateway® clones for rapid follow-up validation

A high-quality human protein microarray with even higher content

The ProtoArray® Human Protein Microarray is an advanced, high-content, functional protein microarray (Figure 1) that enables you to scan thousands of proteins for biochemical interactions in as little as one day. Version 4.0 of the ProtoArray® Human Protein Microarray contains over 8,000 human proteins representing multiple gene families (Table 1), arrayed in duplicate on a 1 inch x 3 inch nitrocellulose-coated slide. The nitrocellulose array provides high protein stability and yields highly sensitive results for protein-protein interaction, kinase substrate identification, and serum profiling studies. In addition, each slide is labeled with a unique bar code for easy tracking.

Maximum protein functionality

The ProtoArray® Human Protein Microarray is designed for maximum protein quality and functionality.[1] ProtoArray® microarray proteins are derived from sequence-validated open reading frames selected from Invitrogen's extensive Ultimate™ ORF Clone Collection.[2] These full-length proteins are expressed as N-terminal GST fusion proteins using a baculovirus-based expression system. All proteins are purified under nondenaturing conditions and printed at +4°C to preserve native structure and proper functionality.

Table 1—ProtoArray® content and protein distribution by class.

Class	No. on array[1]
Protein kinases	378
Transcription factors	188
Membrane proteins	1,085
Nuclear proteins	830
Signal transduction	710
Secreted proteins	100
Cell communication	863
Metabolism	2,166
Cell death	145
Protease/peptidase activity	138

[1] Numbers are based on Gene Ontology databases and are subject to change for a complete list of array proteins, visit www.invitrogen.com/protoarray.

[2] The Ultimate™ Human and Mouse ORF Clone Collection is the first and only clone resource to offer full-insert sequenced clones in a Gateway® vector. To learn more, visit www.invitrogen.com.

Figure 1—Indivi(d)ual image of over 8,000 unique human proteins on a ProtoArray® Human Protein Microarray v4.0.

Rapid and efficient data analysis

ProtoArray® Protein Microarrays can be easily read with most commercially available fluorescent microarray scanners. For a list of compatible scanners, please visit www.invitrogen.com/protoarray. ProtoArray® Prospector v4.0 is a FREE data analysis software tool that evaluates results of control features present on the array and yields data obtained from experimental microarrays. You can download ProtoArray® Prospector v4.0 at www.invitrogen.com/protoarray. Once you have identified your human proteins of interest, you can obtain the corresponding ready-to-use Ultimate™ ORF clones and get started on your downstream validation applications.

Broad range of discovery applications

The ProtoArray® Human Protein Microarray platform enables multiple discovery efforts (Table 2), including:

• Identification of novel protein biomarkers specific for infectious disease, cancers, and autoimmune diseases[2]

• Mapping of protein-protein interactions important in biochemical pathways[3,4]

† Identification of kinase substrates for target discovery and validation[5]

† Antibody specificity profiling for research and therapeutic antibody development[6]

Table 2—The ProtoArray® Human Protein Microarray enables available research applications and experimental flexibility not possible with other approaches.

Application or attribute	Advantage	ProtoArray® Protein Microarray	Peptide arrays	Antibody arrays	Western blotting	Yeast two-hybrid	Mass spectrometry
Involves functional proteins	Biologically relevant results	Yes	No	Yes	No	Yes	Yes
Identified "hits" available as clones	Rapid follow-up validation	Yes, hits available as sequence-verified clones in Gateway® vectors	No	No	No	Yes	No
Low technical complexity	Easy to set up and interpret results	Yes	Yes	Yes	Yes	Yes	No
Protein-protein interaction discovery	Discover and validate drug target pathways	Yes, only 2 days	Limited	No	N/A	Yes, 6 weeks	Yes, 6 weeks
Antibody specificity profiling	Guide selection and development of research and therapeutic antibodies	Yes, detects cross-reactive species via linear and conformational epitopes	Fails to detect conformational epitopes, no specific identification of cross-reactive species	N/A	Fails to detect conformational epitopes, no specific identification of cross-reactive species	N/A	Possible
Biomarker discovery	Identify novel diagnostic markers for disease	Yes, autoimmune biomarkers only	Not demonstrated	Yes	N/A	N/A	Yes, coupled with labeling techniques
Profiling drug binding	Guide drug candidate selection and development	Yes	No	No	N/A	N/A	Yes

ë invitrogen™

Exhibit 7

PROTEIN KINASES 6

The eukaryotic protein kinase superfamily: kinase (catalytic) domain structure and classification[1]

STEVEN K. HANKS* AND TONY HUNTER[2]

*Department of Cell Biology, Vanderbilt University School of Medicine, Nashville, Tennessee 37232, USA; and Molecular Biology and Virology Laboratory, The Salk Institute, San Diego, California 92186, USA

The eukaryotic protein kinases comprise one of the largest superfamilies of homologous proteins and genes. Within this family, there are now hundreds of different members whose sequences are known. Although there is a rich diversity of structures, regulation modes, and substrate specificities among the protein kinases, there are also common structural features. These conserved structural motifs provide clear indications as to how these enzymes manage to transfer the γ-phosphate of a purine nucleotide triphosphate to the hydroxyl groups of their protein substrates. The authors of this review have carried out a monumental task of analyzing and collating the amino acid sequences of all reported protein kinases and defining the conserved structural features that characterize the portion of these proteins that is responsible for their catalytic activity. Comparison of the sequences in the catalytic fragment of the protein kinases has been used to arrange these enzymes in evolutionary trees that group subfamilies of closely related enzymes. It is comforting that the structural relationships that emerge from these trees result in groupings that also reflect related functions. The work presented in this review seems to be an excellent example of the type of analysis that will become indispensable in the coming years, as more and more sequence information become available to biologists as a result of the genome projects.

ABSTRACT   The eukaryotic protein kinases make up a large superfamily of homologous proteins. They are related by virtue of their kinase domains (also known as catalytic domains), which consist of ≈250–300 amino acid residues. The kinase domains that define this group of enzymes contain 12 conserved subdomains that fold into a common catalytic core structure, as revealed by the 3–dimensional structures of several protein–serine kinases. There are two main subdivisions within the superfamily: the protein–serine/threonine kinases and the protein–tyrosine kinases. A classification scheme can be founded on a kinase domain phylogeny, which reveals families of enzymes that have related substrate specificities and modes of regulation.—Hanks, S. K., Hunter, T. The eukaryotic protein kinase superfamily: kinase (catalytic) domain structure and classification. FASEB J. 9, 576–596 (1995)

Key Words: protein–tyrosine kinase • protein–serine kinase • protein phosphorylation • AMP–dependent protein kinase

THE EUKARYOTIC PROTEIN KINASE SUPERFAMILY

One of the largest known protein superfamilies is made up of protein kinases identified largely from eukaryotic sources. (The term superfamily will be used here to distinguish this broad collection of enzymes from smaller, more closely related subsets that have been commonly referred to as families). These enzymes use the γ-phosphate of ATP (or GTP) to generate phosphate monoesters using protein alcohol groups (on Ser and Thr) and/or protein phenolic groups (on Tyr) as phosphate acceptors. The protein kinases are related by virtue of their homologous kinase domains (also known as catalytic domains), which consist of ~250–300 amino acid residues (reviewed in refs 1–3; and see below). During the past 15 years, previously unrecognized members of the eukaryotic protein kinase superfamily have been uncovered at an exponentially increasing rate and currently appear in the literature almost weekly. This pace of discovery can be attributed to the past development of molecular cloning and sequencing technologies and, more recently, to the advent of the polymerase chain reaction (PCR),[3] which facilitated the use of homology–based cloning strategies. Consequently, about 200 different superfamily members (products of distinct paralogous genes) had been recognized from mammalian sources alone! The prediction made several years ago (4) that the mammalian genome contains about 1000 protein kinase genes (roughly 1% of all genes) would still appear to be within reason, and may even be an underestimate (5).

In addition to mammals and other vertebrates, eukaryotic protein kinase superfamily members have been identified and characterized from a wide range of other animal phyla as well as from plants, fungi, and protozoans. Hence, the protein kinase progenitor gene can be traced back to a time before the evolutionary separation of the major eukaryotic kingdoms. The identification of eukaryotic–like protein kinase genes in prokaryotes (6, 7) raises the possibility that the protein kinase progenitor gene might have arisen before the divergence of prokaryotes and eukaryotes (see below). Studies of the budding and fission yeasts, Saccharomyces cerevisiae and Schizosaccharomyces pombe, have been particularly fruitful in the recognition of new protein kinases. In these geneti-

cally tractable organisms, the powerful approach of mutant isolation and cloning by complementation has netted dozens of protein kinase genes required for numerous aspects of cell function (8). In many cases, vertebrate counterparts have now been found for these genes, leading to a growing awareness that protein phosphorylation pathways that regulate basic aspects of cell physiology have been maintained throughout the course of eukaryotic evolution.

Even though the overwhelming majority of protein kinases identified from eukaryotic sources belong to this superfamily, a small but growing number of such enzymes do not qualify as superfamily members. Most of these are related to the prokaryotic protein-histidine kinase family (see below), which forms the sensor components of two-component signal transduction systems (9). Included in this category are a putative ethylene receptor encoded by the flowering plant ETR1 gene (10), the product of the budding yeast SLN1 gene (11, 12) thought to be involved in relaying nutrient information to elements controlling cell growth and division, the mitochondrial branched-chain α-ketoacid dehydrogenase kinase (13), and the mitochondrial pyruvate dehydrogenase kinase (14). In prokaryotes, protein-histidine kinases phosphorylate aspartates in their target proteins, but except for the two dehydrogenase kinases that phosphorylate serine, the acceptor specificities of most of the eukaryotic protein kinases of this type are not known. In addition to these protein kinases, the Bcr protein encoded by the breakpoint cluster region gene involved in the Philadelphia chromosome translocation (15) and the A6 kinase isolated by expression cloning using an anti-phosphotyrosine antibody (16) have kinase domains unrelated to any known eukaryotic or prokaryotic kinase. In addition, true protein-histidine kinases are known in eukaryotes. One such enzyme has been extensively characterized from budding yeast but not yet molecularly cloned (17), and so it is not clear whether this enzyme will belong to the protein kinase superfamily or use a novel structural principle for phosphotransfer.

What about the prokaryotes? It has been known for years that protein phosphorylation events play key regulatory roles in numerous bacterial cell processes including chemotaxis, bacteriophage infection, nutrient uptake, and gene transcription (reviewed in refs 18, 19). The bacterial protein kinases have been divided into three general classes (20): 1) protein-histidine kinases such as those functioning in two-component sensory regulatory systems (strictly speaking, these are protein-aspartyl kinases, because autophosphorylation on His is an intermediary step in phosphotransfer to an aspartate in the response-regulator protein) (9); 2) phosphotransferases such as those of the phosphoenol pyruvate-dependent phosphotransferase system involved in sugar uptake (21); and 3) protein-serine kinases such as isocitrate dehydrogenase kinase/phosphatase (22). Amino acid sequences have been determined for members of each class, and all are unrelated to the eukaryotic protein kinase superfamily.

Recently, however, true homologs of the eukaryotic protein kinases have been identified from two species of bacteria, Yersinia pseudotuberculosis (7) and Myxococcus xanthus (6, 23). Are these special cases, or the first examples of many such genes in prokaryotes? The eukaryotic-like protein kinase YpkA from the pathogenic enterobacteria Y. pseudotuberculosis is encoded by a plasmid essential for

the virulence of this infectious organism. In addition to YpkA, at least two other proteins encoded by genes residing on the virulence plasmid exhibit high similarity to eukaryotic proteins. Thus, it seems likely that the virulence plasmid genes were transduced from a eukaryotic host by horizontal transfer. The myxobacterium M. xanthus presents a different and perhaps more intriguing picture. Application of the PCR homology-based cloning strategy revealed that at least eight genes encoding members of the eukaryotic protein kinase superfamily are present in the genome of this species (23). The myxobacteria are unusual prokaryotes in that they undergo a complex developmental cycle upon nutrient depletion, much like that of the eukaryotic slime mold Dictyostelium. Given that protein kinases are commonly involved in regulating growth and differentiation of eukaryotic cells, it is attractive to speculate that the eukaryotic-like protein kinases in M. xanthus are specifically involved in regulating their developmental cycle. Indeed, one of these kinases, Pkn1, was shown to be required for proper fruiting body formation. The same could be true for the eukaryotic-like protein kinase PknA from Anabena (24). In keeping with this idea, neither the PCR approach applied to Escherichia coli (23) nor extensive sequencing of the E. coli genome (now 30% complete) has yielded eukaryotic-like protein kinases. Hence, genes encoding members of the eukaryotic protein kinase superfamily may be present only in bacteria that can undergo a developmental cycle. However, unpublished reports of eukaryotic-like protein kinases in Streptomyces coelicolor, and in three species of Methanococcus, suggest that such genes are more widely expressed among prokaryotes, and potentially these genes represent the ancestors for the entire eukaryotic protein kinase superfamily.

THE HOMOLOGOUS KINASE DOMAINS

The kinase domains of eukaryotic protein kinases impart the catalytic activity. Three separate roles can be ascribed to the kinase domains: 1) binding and orientation of the ATP (or GTP) phosphate donor as a complex with divalent cation (usually Mg^{2+} or Mn^{2+}); 2) binding and orientation of the protein (or peptide) substrate; and 3) transfer of the γ-phosphate from ATP (or GTP) to the acceptor hydroxyl residue (Ser, Thr, or Tyr) of the protein substrate.

Conserved features of primary structure

The total number of distinct kinase domain amino acid sequences available is now approaching 400 (Table 1). Included in this total are the vertebrate enzymes encoded by distinct paralogous genes, their presumed functional homologs from invertebrates and simpler organisms (encoded by orthologous genes), and those identified from lower organisms and plants for which vertebrate equivalents have not been found. Conserved features of kinase domain primary structure have previously been identified through an inspection of multiple amino acid sequence alignments (1-3) . The large number of sequences now available precludes showing an alignment containing all known kinase domains. Thus, in Fig. 1 only 60 different kinase domain sequences are aligned. These are drawn, however, from the widest possible sampling of the superfamily and thus provide a good representation of the

Table 1. *Eukaryotic protein kinase superfamily classification.*

A-C-G Group
  AGC-I. Cyclic nucleotide-regulated protein kinase family
    A. Cyclic AMP-dependent protein kinase (PKA) subfamily
      vertebrate:

1. PKA-Cα:	PKA catalytic subunit, alpha-form	
2. PKA-Cβ:	PKA catalytic subunit, beta-form	
3. PKA-Cγ:	PKA catalytic subunit, gamma-form	

      Drosophila melanogaster:

1. DmPKA-C0:	PKA catalytic subunit, C0 form
2. DmPKA-C1:	PKA catalytic subunit, C1 form
3. DmPKA-C2:	PKA catalytic subunit, C2 form

      Caenorhabditis elegans:
        1. CePKA: — PKA catalytic subunit homolog
      Saccharomyces cerevisiae:
        1. ScPKA-Tpk1: — PKA catalytic subunit homolog, type 1
      Schizosaccharomyces pombe:
        1. SpPKA1: — PKA catalytic subunit homolog
      Dictyostelium discoideum:
        1. DdPKA: — PKA catalytic subunit
      Aplysia californica:
        1. AplC: — PKA catalytic subunit homolog
        2. Sak: — "Spermatozoon-associated kinase"
    B. Cyclic GMP-dependent protein kinase (PKG) subfamily
      vertebrate:
        1. PKG-I: — PKG, type I
      * 2. PKG-II: — PKG, type II
      Drosophila melanogaster:
        1. DmPKG-G1: — PKG homolog, type 1
        2. DmPKG-G2: — PKG homolog, type 2
    C. Others
      Dictyostelium discoideum:
        1. DdPK1: — PKA homolog

  AGC-II. Diacylglycerol-activated/phospholipid-dependent protein kinase C (PKC) family
    A. "Conventional" (Ca^+-dependent) protein kinase C (cPKC) subfamily
      vertebrate:
        1. cPKCα: — Protein Kinase C, alpha-form
        2. cPKCβ: — Protein Kinase C, beta-form
        3. cPKCγ: — Protein Kinase C, gamma-form
      Drosophila melanogaster:
        1. DmPKC-53Ebr: — PKC homolog expressed in brain, locus 53E
        2. DmPKC-53Eey: — PKC homolog expressed in eye, locus 53E
      Aplysia californica:
        1. Apl-I: — PKC homolog, type I
    B. "Novel" (Ca^+-independent) Protein Kinase C (nPKC) subfamily
      vertebrate:
        1. nPKCδ: — Protein Kinase C, delta-form
        2. nPKCε: — Protein Kinase C, epsilon-form
        3. nPKCη: — Protein Kinase C, eta-form
        4. nPKCθ: — Protein Kinase C, theta-form
      Drosophila melanogaster:
        1. DmPKC-98F: — PKC homolog, locus 98F
      Aplysia californica:
        1. Apl-II: — PKC homolog, type II
      Caenorhabditis elegans:
        1. CePKC: — PKC homolog, product of *tpa-1* gene
      * 2. CePKC1B: — PKC homolog expressed in neurons and interneurons
      Dictyostelium discoideum:
      * 1. DdMHCK: — PKC homolog
      Saccharomyces cerevisiae:
        1. ScPKA1: — PKC homolog, product of *PKC1* gene
      * 2. ScPKA2: — PKC homolog, product of *PKC2* gene
      Schizosaccharomyces pombe:
        1. Pck1: — "Pombe C-kinase", type 1
        2. Pck2: — "Pombe C-kinase", type 2
    C. "Atypical" Protein Kinase C (aPKC) subfamily
      vertebrate:
        1. aPKCζ: — Protein Kinase C, zeta-form
      * 2. aPKCι: — Protein Kinase C, iota-form
      * 4. aPKCμ: — Protein Kinase C, mu-form

Table 1. *(continued)*.

D. Others
   vertebrate:
   * 1. PKN: Protein kinase with PKC-related catalytic domain

AGC-III. Related to PKA and PKC (RAC) family
   vertebrate:
      1. RAC-α: RAC, alpha-form; cellular homolog of v-Akt oncoprotein
      2. RAC-β: RAC, beta-form
   Drosophila:
      1. DmRAC: RAC homolog
   Caenorhabditis elegans:
   * 1. CeRAC: RAC homolog

ACG-IV. Family of kinasese that phosphorylate G protein-coupled receptors
   vertebrate:
      1. βARK1:   β-adrenergic receptor kinase, type 1
      2. βARK2:   β-adrenergic receptor kinase, type 2
      3. RhK: Rhodopsin kinase
   * 4. IT11:   G-protein-coupled receptor kinase homolog
   * 5. GRK5:   G-protein-coupled receptor kinase, type 5
   * 6. GRK6:   G-protein-coupled receptor kinase, type 6
   Drosophila melanogaster:
      1. DmGPRK1:   Drosophila G-protein-coupled receptor kinase, type 1
      2. DmGPRK2:   Drosophila G-protein-coupled receptor kinase, type 2

AGC-V. Family of budding yeast AGC-related kinases
   Saccharomyces cerevisiae:
      1. Sch9:   Suppressor of defects in cAMP effector pathway
      2. Ykr2:   AGC-related kinase
      3. Ypk1:   AGC-related kinase

AGC-VI. Family of kinases that phosphorylate ribosomal S6 protein
   vertebrate:
      1. S6K: 70 kDa S6 kinase with single catalytic domain
      2. RSK1(Nt):   90 kDA S6 kinase, type 1
      3. RSK2(Nt):   90 kDA S6 kinase, type 2
      [Note: The RSK enzymes have two distinct catalytic domains. The Nt-domain is closely related to S6K, whereas the Ct-domain is most closely related to phosphorylase kinase]

AGC-VII. Budding yeast Dbf2/20 Family
   Saccharomyces cerevisiae:
      1. Dbf2:   Product of gene periodically expressed in cell cycle
      2. Dbf20:   Close relative of DBF2 not under cell cycle control

AG-VIII. Flowering plant "PVPK1 Family" of protein kinase homologs
   Phylum Angiospermophyta (Kingdom Plantae):
      1. PvK1:   Bean protein kinase homolog
      2. OsG11A: Rice protein kinase homolog
      3. ZmPPK: Maize protein kinase homolog
      4. AtPK5:   Arabidopsis protein kinase homolog
      5. AtPK7:   Arabidopsis protein kinase homolog
      6. AtPK64:   Arabidopsis protein kinase homolog
      7. PsPK5:   Pea protein kinase homolog

Other AGC-related kinases
   vertebrate:
      1. DMPK: "Myotonic Dystrophy Protein Kinase"
      2. Sgk: "Serum and glucocortocoid regulated kinase"
   * 3. Mast205:   Spermatid "Microtubule-associated serine/threonine kinase"
   Neurospora crassa:
      1. NcCot-1:   Product of gene required for normal colonial growth
   Dictyostelium discoideum:
      1. Ddk2:   Product of developmentally-regulated gene
   Saccharomyces cerevisiae:
      1. ScSpk1:   Dual-specificity kinase
   Phylum Angiospermophyta (Kingdom Plantae):
   * 1. Atpk1:   Arabidopsis protein kinase

CaMK Group
   CaMK-I. Family of kinases regulated by Ca²⁺/Calmodulin, and close relatives
   A. Subfamily including "Multifunctional" Ca^+/Calmodulin Kinases (CaMKs)
   vertebrate:
      1. CaMK1:   CaMK, type I
      2. CaMK2α: CaMK, type II, alpha subunit
      3. CaMK2β: CaMK, type II, beta subunit
      4. CaMK2γ: CaMK, type II, gamma subunit
      5. CaMK2δ: CaMK, type II, delta subunit
   * 6. EF2K: Elongation Factor-2 Kinase or CaMK type III
      7. CaMK4:   CaMK, type IV

Table 1. *(continued).*

Drosophila melanogaster:
    1. DmCaMK2:   CaMK-II homolog
Saccharomyces cerevisiae:
    1. ScCaMK2-1:   CaMK-II homolog, product of *CMK1* gene
    2. ScCaMK2-2:   CaMK-II homolog, product of *CMK2* gene
Aspergillus nidulans:
    1. AnCaMK2:   CaMK-II homolog

B. Subfamily including phosphorylase kinases
   vertebrate:
    1. PhK-γM: Skeletal muscle phosphorylase kinase catalytic subunit
    2. PhK-γT: Male germ cell phosphorylase kinase catalytic subunit
    3. RSK1(Ct):   90 kDa S6 kinase, type 1; C-terminal catalytic domain
    4. RSK2(Ct):   90 kDa S6 kinase, type 2; C-terminal catalytic domain

C. Subfamily including myosin light chain kinases
   vertebrate:
    1. skMLCK: Skeletal muscle MLCK (rabbit)
    2. smMLCK: Smooth muscle MLCK (rabbit)
    3. Titin: Huge protein implicated in skeletal muscle development
   Caenorhabditis elegans:
    1. Twn: "Twitchin" protein involved in muscle contraction or development
   Dictyostelium discoideum:
    1. DdMLCK: Slime mold myosin light chain kinase

D. Subfamily of plant kinases with intrinsic calmodulin-like domain
   Phylum Angiospermophyta (Kingdom Plantae):
    1. CDPK: Soybean Ca^+-regulated kinase with intrinsic CaM-like domain
    2. AtAK1:   Arabidopsis CDPK homolog
  * 3. OsSpk: Rice CDPK homolog
  * 4. DcPk431:   Carrot CDPK homolog

E. Subfamily of plant kinases with highly acidic domain
   Phylum Angiospermophyta (Kingdom Plantae):
  * 1. ASK1:   Arabidopsis protein kinase homolog with highly acidic idomain
  * 2. ASK2:   Arabidopsis protein kinase homolog with highly acidic domain

F. Other CaMK-related kinases
   vertebrate:
    1. PskH1:   Putative protein-serine kinase
  * 2. MAPKAP2:   "MAP Kinase-Activated Protein Kinase 2"
   Saccharomyces cerevisiae:
    1. Mre4:   Protein required for meiotic recombination
  * 2. Dun1:   Protein required for DNA damage-inducible gene expression
  * 3. Rck1:   "Radiation sensitivity complementing kinase, type 1"
  * 4. Rck2:   "Radiation sensitivity complementing kinase, type 2"

CaMK-II. Snf1/AMPK family
   vertebrate:
  * 1: AMPK: "AMP-Activated Protein Kinase"
    2: p78:   Protein lost in carcinomas of human pancreas
   Saccharomyces cerevisiae:
    1. Snf1:   Kinase essential for release from glucose repression
    2. Kin1:   Protein kinase with N-terminal catalytic domain
    3. Kin2:   Close relative of KIN1
    4. Ycl24:   Protein kinase homolog on chromosome III
  * 5. Ycl453:   Protein kinase homolog on chromosome XI
   Schizosaccharomyces pombe:
    1. SpKin1:   Product of gene important for growth polarity
    2. Nim1:   Inducer of mitosis
   Phylum Angiospermophyta (Kingdom Plantae):
    1. PSnf1-RKIN1:   Rye putative protein kinase that complements yeast *snf1* polarity
    2. PSnf1-AKIN10:   Arabidopsis putative protein kinase related to SNF1
    3. PSnf1-BKIN12:   Barley protein related to SNF1
  * 4. PKABA1:   Wheat kinase induced by abscisic acid
  * 5. WPK4:   Wheat kinase homolog regulated by light and nutrients
  * 6. NPK5:   Tobacco Snf1 homolog, activates *SUC2* gene expression

Other CaMK Group Kinases
   Plasmodium falciparum (malarial parasite):
    1. PfCPK: Ca^+-regulated kinase with intrinsic CaM-like domain
    2. PfPK2:   Putative protein kinase

C-M-C-C Group
  CMGC-I. Family of cyclin-dependent kinases (CDKs) and other close relatives
   vertebrate:
    1. Cdc2:   Inducer of mitosis; functional homolog of yeast cdc2+/CDC28 kinases (Cdk1)
    2. Cdk2:   Type 2 cyclin-dependent kinase
    3. Cdk3:   Type 3 cyclin-dependent kinase
    4. Cdk4:   Type 4 cyclin-dependent kinase
    5. Cdk5:   Type 5 cyclin-dependent kinase

Table 1. (continued).

6. Cdk6:	Type 6 cyclin-dependent kinase
7. PCTAIRE1:	Cdc2-related protein
8. PCTAIRE2:	Cdc2-related protein
9. PCTAIRE3:	Cdc2-related protein
10. Mo15:	"Cdk-activating kinase"; Negative regulator of meiosis (CAK)

Drosophila melanogaster:

1. DmCdc2:	Functional homolog of yeast cdc2+/CDC28 kinases
2. DmCdc2c:	Cdc2-cognate protein; Cdk2 homolog

Dictyostelium discoideum:

1. DdCdc2:	Functional homolog of yeast cdc2+/CDC28 kinases
2. DdPRK:	"Cdc2-related PCTAIRE Kinase"

Aspergillus nidulans:

1. NIMXcdc2:	Cdc2-related gene product

Plasmodium falciparum:

1. PfPK5:	Cdc2-related protein from human malarial parasite

Entamoeba histolytica:

1. EhC2R:	Cdc2-related protein

Crithidia fasciculata:

1. CfCdc2R:	Cdc2-related protein

Leishmania mexicana:

* 1. LmCRK1:	"Cdc2-Related Kinase"

Saccharomyces cerevisiae:

1. Cdc28:	"Cell-division-cycle" gene product
2. Pho85:	Negative regulator of the PHO system and cell cycle regulator
3. Kin28:	CDC28-related protein

Schizosaccharomyces pombe:

1. SpCdc2:	"Cell-division-cycle" gene product

Histoplasma capsulatum:

* 1. HcCdc2:	Cdc2 homolog from dimorphic fungus

Phylum Angiospermophyta (Kingdom Plantae):

1. Pcdc2:	Flowering plant Cdc2 homolog othat complements yeast mutants
* 2. MsCdc2B:	Alfalfa Cdc2 cognate gene products that complements G1/S transition
3. OsC2R:	More distantly related Cdc2 homolog from rice

CMGC-II. Erk(MAP kinase) family

vertebrate:

1. Erk1:	"Extracellular signal-regulated kinase", type 1 (p44 MAP kinase)
2. Erk2:	"Extracellular signal-regulated kinase", type 2 (p42 MAP kinase)
3. Erk3:	Somewhat distant relative of the Erk/MAP kinases
* 4. p63MAPK:	Another more distant relative of the Erk/MAP kinases
* 5. SAPK-α:	"Stress-activated protein kinase, type alpha" (JNK2)
* 6. SAPK-β:	"Stress-activated protein kinase, type beta"
* 7. SAPK-γ/Jnk1:	"Stress-activated protein kinase, type gamma" or "Jun N-terminal Kinase"
* 8. p38:	HOG1-related protein (MPK2)

Drosophila melanogaster:

1. DmErkA:	Homolog of Erk/MAP kinases; product of *rolled* gene

Caenorhabditis elegans:

* 1. Sur1:	Erk/MAP kinase homolog

Saccharomyces cerevisiae:

1. Kss1:	Suppressor of *sst2* mutant, overcomes growth arrest
2. Fus3:	Product of gene required for growth and mating
3. Slt2:	Product of gene complementing *byt2* mutants (MPK1)
* 4. Hog1:	Product of gene required for osmoregulation

Schizosaccharomyces pombe:

1: Spk1:	Product of gene that confers drug resistance to staurosporine, a PK inhibitor

Phylum Deuteromycota (Kingdom Fungi):

1. CaErk1:	Protein that interferes with mating factor-induced cell cycle arrest

Trypanosoma brucei (Phylum Zoomastigina, Kingdom Protoctista):

* 1. KFR1:	"KSS1- and FUS3-related" gene product

Phylum Angiospermophyta (Kingdom Plantae):

1. PErk:	Flowering plant Erk/MAP kinase homologs (7 distinct homologs identified in Arabidopsis)

CMGC-III. Glycogen synthase kinase 3 (GSK3) family

vertebrate:

1. GSK3α:	Glycogen synthase kinase 3, α-form
2. GSK3β:	Glycogen synthase kinase 3, β-form

Drosophila melanogaster:

1. Sgg:	Product of *shaggy/zeste-white 3* gene

Saccharomyces cerevisiae:

1. Mck1:	"Meiosis and centromere regulatory kinase"
* 2. ScGSK3	Protein closely related to MCK1
* 3. Mds1:	Dosage suppressor of mck1 mutant

Dictyostelium discoideum:

* 1. DdGSK3:	Glycogen synthase kinase 3 homolog

Phylum Angiospermophyta (Kingdom Plantae):

* 1. ASK-α:	"Arabidopsis shaggy-related protein kinase", type alpha
* 2. ASK-γ:	"Arabidopsis shaggy-related protein kinase", type gamma

Table 1. *(continued).*

vertebrate:
   1. CK2α: Casein kinase II, alpha subunit
   1. CK2α':   Casein kinase II, alpha-prime subunit
Drosophila melanogaster:
   1. DmCK2:   Casein kinase II homolog
Caenorhabditis elegans:
   1. CeCK2:   Casein kinase II homolog
Theileria parva (a protozoan parasite):
   1. TpCK2:   Casein kinase II α-subunit homolog
Dictyostelium discoideum:
   1. DdCK2:   Casein kinase II, α-subunit
Saccharomyces cerevisiae:
   1. ScCK2α: Casein kinase II, alpha subunit
   2. ScCK2α':   Casein kinase II, alpha-prime subunit
Schizosaccharomyces pombe:
*   1. SpCka1:   Casein kinase II, α-subunit homolog (Orb5)
Phylum Angiospermophyta (Kingdom Plantae):
   1. ZmCK2:   Flowering plant casein kinase II, α-subunit homolog

CMGC-IV. Clk family
vertebrate:
   1. Clk: "Cdc-like kinase"
*   2. Srpk1:   Kinase that regulates intracellular localization of splicing factors
   3. PskG1:   Putative protein kinase
   4. PskH2:   Putative protein kinase
Drosophila melanogaster:
*   1. Doa: Kinase encoded by "Darkener of Apricot" locus
Saccharomyces cerevisiae:
   1. Yak1:   Suppressor of RAS mutant
   2. Kns1:   Nonessential protein kinase homolog
Schizosaccharomyces pombe:
   1. Dsk1:   Dis1-suppressing protein kinase implicated in mitotic control
*   2. Prp4:   Pre-mRNA processing gene product; lacks subdomains X-XI

Other CMGC Group kinases
vertebrate:
   1. Mak: "Male germ cell-associated kinase"
   2. Ched: "Cholinesterase-related cell division controller"
   3. PITSLRE: Galactosyltransferase-associated kinase
   4. KKIALRE: Cdc2-related protein
*   5. PITALRE: Cdc2-related kinase
*   6. PISSLRE: Cdc2-related kinase
Saccharomyces cerevisiae:
   1. Sme1:   Product of gene essential for start of meiosis
   2. Sgv1:   Kinase required for G-protein-mediated adaptive response to pheromone
   3. Ctk1:   Product of gene required for normal growth
Phylum Angiospermophyta (Kingdom Plantae):
*   1. Mhk: Arabidopsis thaliana "Mak homologous kinase"

Conventional Protein-Tyrosine Kinase Group (I-X: Non-membrane-spanning; XI-XXIII: Membrane-spanning)
PTK-I. Src family
vertebrate:
   1. Src: Cellular homolog of Rous sarcoma virus oncoprotein
   2. Yes: Cellular homolog of Yamaguchi 73 sarcoma virus oncoprotein
   3. Yrk: Yes-related kinase
   4. Fyn: Protein related to Fgr and Yes
   5. Fgr: Cellular homolog of Gardner-Rasheed sarcoma virus oncoprotein
   6. Lyn: Protein related to Fgr and Yes
   7. Hck: Hematopoietic cell protein-tyrosine kinase
   8. Lck: Lymphoid T-cell protein-tyrosine kinase
   9. Blk: Lymphoid B-cell protein-tyrosine kinase
*   10. Frk: Fyn-related kinase
*   11. Rak: STK-related kinase
*   12. Fyk: "Fyn and Yes-related kinase" from electric ray
Drosophila melanogaster:
   1. DmSrc: Src homolog, polytene locus 64B
Dugesiai (Girardia) tigrina (Phylum Platyhelminthes):
*   1. DtSpk-1:   "Src-like planarian kinase"
Hydra vulgaris (Phylum Cnidaria):
   1. Stk: Src-related protein
Spongilla lacustris (Phylum Porifera):
   1. Srk1-4:   Four distinct Src-related kinases

PTK-II. Brk family
vertebrate:
*   1. Brk: Protein-tyrosine kinase expressed in human breast tumors

Table 1. *(continued).*

PTK-III. Tec family
vertebrate:
    1. Tec: "Tyrosine kinase expressed in hepatocellular carcinoma"
    2. Emt: "Expressed mainly in T-cells" kinase (Itk, Tsk)
    3. Btk: "Bruton's agammaglobulinaemia tyrosine kinase" (Emb)
*   4. Txk: Tec-related protein-tyrosine kinase
Drosophila melanogaster:
    1. DmTec: Tec homolog, polytene locus 28C

PTK-IV. Csk family
vertebrate:
    1. Csk: "C terminal Src Kinase"; negative regulator of Src
*   2. MatK: "Megakaryocyte-associated Tyr-kinase" (Hyl, Lsk, Ctk, Ntk)

PTK-V. Fes(Fps) family
vertebrate:
    1. Fes/Fps: Cellular homolog of feline and avian sarcoma viruses
    2. Fer: "Fes/Fps-related" kinase
Drosophila melanogaster:
    1. DmFer: Fer-related protein

PTK-VI. Abl family
vertebrate:
    1. Abl: Cellular homolog of Abelson murine leukemia virus
    2. Arg: "Abl-related gene" product
Drosophila melanogaster:
    1. DmAbl: Abl-related protein
Caenorhabditis elegans:
    1. CeAbl: Nematode Abl-related protein

PTK-VII. Syk/Zap70 family
vertebrate:
    1. Syk: "Spleen tyrosine kinase"
    2. Zap70:   T-cell receptor "zeta chain-associated protein of 70 kDa"
Hydra vulgaris (Phylum Cnidaria):
*   1. Htk16:   Syk/Zap70-related

PTK-VIII. Jak family
vertebrate:
    1. Tyk2:   Transducer of interferon α/β signals
    2. Jak1:   "Janus kinase", type 1
    3. Jak2:   "Janus kinase", type 2
*   4. Jak3:   "Janus kinase", type 3
Drosophila melanogaster:
*   1. Hop: Product of hopscotch gene required for establishing segmental body plan

PTK-IX. Ack
vertebrate:
*   1. Ack: "CDC42Hs-associated kinase"

PTK-X. Fak
vertebrate:
    1. Fak: "Focal adhesion kinase"

PTK-XI. Epidermal growth factor receptor family
vertebrate:
    1. EGFR: Epidermal growth factor receptor
    2. ErbB2:   Cell homolog of oncogene activated in ENU-induced rat neuroblastoma (Neu, HER2)
    3. ErbB3:   Receptor tyrosine kinase related to EGFR (HER3)
    4. ErbB4:   Receptor tyrosine kinase related to EGFR (Tyro2)
Drosophila melanogaster:
    1. DER: Homolog of EGF receptor
Caenorhabditis elegans:
    1. LET-23:   Product of gene required for normal vulval development
Schistosoma mansoni (Phylum Platyhelminthes):
    1. SER: EGF receptor homolog

PTK-XII. Eph/Elk/Eck receptor family
vertebrate:
    1. Eph: Kinase detected in "erythropoietin-producing hepatoma"
    2. Eck: "Epithelial cell linase"
    3. Eek: Eph/Elk-related protein-tyrosine kinase
    4. Hek: Eph/Elk related protein-tyrosine kinase (Cek4)
    5. Sek: "Segmentally-expressed kinase"
    6. Elk: "Eph-like kinase" detected in brain
*   7. Hek2:   "Human embryo kinase" type 2 (Cek10)
*   8. Htk: "Hepatoma transmembrane kinase"
    9. Cek5/Nuk: "Chicken embryo kinase 5"/"Neural kinase"
*   10. Ehk1:   "Eph homology kinase-1" (Cek7)
*   11. Ehk2:   "Eph homology kinase-2"
*   12. Myk1:   "Mammary-derived tyrosine kinase, type 1"

Table 1. *(continued).*

•	13. Myk2:	"Mammary-derived tyrosine kinase, type 2"
•	14. Cek9:	"Chicken embryo kinase 9"
•	15. Pag:	"Pagliaccio" Xenopus protein expression in neural crest and neural tissues
•	16. Rtk1:	Zebrafish Eph/Elk-related protein-tyrosine kinase
•	17. Rtk2:	Zebrafish Eph/Elk-related protein-tyrosine kinase
•	18. Rtk3:	Zebrafish Eph/Elk-related protein-tyrosine kinase

PTK-XIII. Axl family
vertebrate:
    1. Axl: "Anexelekto" (Gr. "uncontrolled") tyrosine kinase (UFO, Ark)
    2. Eyk: Cellular homolog of RPL30 avian oncoprotein (c-Ryk)
•   3. Brt/Sky/Tif/Rse: "Brain tyrosine kinase"/"Sea related protein tyrosine kinase"/"Tyrosine kinase with Ig-like and FN-III-like domains"/"Receptor sectaris" (Tyro3)

PTK-XIV. Tie/Tek family
vertebrate:
    1. Tie: "Tyrosine kinase with Ig and EGF homology"
    2. Tek: "Tunica interna endothelial cell kinase" (TIE2)

PTK-XV. Platelet-derived growth factor receptor family
  A. Subfamily witih 5 Ig-like extracellular domains
  vertebrate:
    1. PDGFRα: Platelet-derived growth factor receptor, type alpha
    2. PDGFRβ: Platelet-derived growth factor receptor, type beta
    3. CSF1R: Colony-stimulating factor-1 receptor (c-Fms)
    4. Kit: Steel growth factor receptor
    5. Flk2:   "Fetal liver kinase-2" (Flt3)
  B. Subfamily with 7 Ig-like extracellular domains
  vertebrate:
    1. Flt1:   "Fms-like tyrosine kinase", type 1
    2. Flt4:   "Fms-like tyrosine kinase", type 4
    3. Flk1:   "Fetal liver kinase-1" (KDR)

PTK-XVI. Fibroblast growth factor receptor family
vertebrate:
    1. FGFR1:   Fibroblast growth factor receptor, type 1 (Flg, Cek1)
    2. FGFR2:   Fibroblast growth factor receptor, type 2 (Bek, K-SAM, Cek3)
    3. FGFR3:   Fibroblast growth factor receptor, type 3
    4. FGFR4:   Fibroblast growth factor receptor, type 4
  Drosophila melanogaster:
    1. DmFGFR1:   Fibroblast growth factor receptor homolog, type 1
•   2. DmFGFR2:   Fibroblast growth factor receptor homolog, type 2

PTK-XVII. Insulin receptor family
vertebrate:
    1. InsR: Insulin receptor
    2. IGF1R: Insulin-like growth factor receptor
    3. IRR: Insulin receptor-related protein
  Drosophila melanogaster:
    1. DmInsR: Homolog of insulin receptor

PTK-XVIII. Ltk/Alk family
vertebrate:
    1. Ltk: "Leukocyte tyrosine kinase
•   2. Alk: "Anaplastic lymphoma kinase

PTK-XIX. Ros/Sev family
vertebrate:
    1. Ros: Cellular homolog of UR2 avian sarcoma virus oncoprotein
  Drosophila melanogaster:
    1. Sev: Product of *sevenless* gene required for R7 photoreceptor cell development

PTK-XX. Trk/Ror family
vertebrate:
    1. Trk: High molecular weight nerve growth factor receptor
    2. TrkB: Receptor for nrain-derived neurotrophic factor and neurotrophin-4/5
    3. TrkC: Trk-related protein; receptor for neurotrophin-3
    4. Ror1:   "Ror" putative receptor, type 1
    5. Ror2:   "Ror" putative receptor, type 2
    6. TcRTK: Trk-related receptor (electric ray)
  Drosophila melanogaster:
•   1. Dror: Putative neurotrophic receptor

PTK-XXI. Ddr/Tkt family
•   1. Ddr: "Discoidin Domain Receptor" (TrkE, CAK, NEP, Ptk3)
•   2. Tkt: "Tyrosine Kinase Related to Trk" (Tyro 10)

Table 1. *(continued).*

PTK-XXII. Hepatocyte growth factor receptor family
vertebrate:
    1. HGFR: Hepatocyte growth factor receptor (MET)
    2. Sea: Cellular homolog of S13 avian erythroleukemia virus oncoprotein
    3. Ron: "Recepteur d'Origine Nantaise"
  • 4. Stk: "Stem cell-derived tyrosine kinase"

PTK-XXIII. Nematode Kin15/16 family
Caenorhabditis elegans:
    1. CeKin15:   PTK expressed during hypodermal development
    2. CeKin16:   PTK expressed during hypodermal development

Other membrane-spanning protein-tyrosine kinases (each with no close relatives)
vertebrate:
    1. Ret: Normal homolog of oncoprotein activated by recombination
    2. Klg: "Kinase-like gene" product
  • 3. Nyk/Ryk: "Novel tyrosine kinase-related protein" (VIK, Mrk, Nbtk1)
Drosophila melanogaster:
    1. Torso: Product of *torso* gene required for embryonic anterior/posterior determination
    2. DmTrk: Distant relative of the mammalian trk gene
Marine sponge (Geodia cydonium):
  • 1. GCTK: Putative receptor PTK

Other protein kinase families (not falling into m ajor groups)
O-I. Polo family
vertebrate:
    1. Plk: "Polo-like kinase"
    2. Snk: "Serum-inducible kinase"
  • 3. Sak: Polo-related kinase isolated in screen for genes regulating sialylation
Drosophila melanogaster:
    1. Polo: Protein kinase homolog required for mitosis
Saccharomyces cerevisiae:
    1. Cdc5:   Product of gene required for cell cycle progression

O-II. MEK/STE7 family
vertebrate:
    1. MEK1:   "MAP ERK Kinase", type 1
    2. MEK2:   "MAP ERK Kinase", type 2
Drosophila melanogaster:
    1. Dsor1:
Saccharomyces cerevisiae:
    1. Ste7:   Kinase required for haploid-specific gene expression
    2. Pbs2:   Kinase required for antibiotic drug resistance
    3. Mkk1:   "MAP Kinase Kinase", type 1 (suppresses lysis defect of pkc1 mutant)
    4. Mkk2:   "MAP Kinase Kinase", type 2 (suppresses lysis defect of pkc1 mutant)
Schizosaccharomyces pombe:
    1. Byr1:   Kinase that suppresses ras1-mutant sporulation defect
    2. Wis1:   Suppressor of cdc phenotype in triple mutant *cdc25/wee1/win1* strains

O-III. MEKK/Ste11 family
vertebrate:
  • 1. MEKK: "MEK Kinase"
Saccharomyces cerevisiae:
    1. Ste11:   Protein required for cell-type-specific transcription
    2. Bck1:   "Bypass of C kinase" kinase
Schizosaccharomyces pombe:
    1. Byr2:   Product of gene required for pheromone signal transduction
Phylum Angiospermophyta (Kingdom Plantae):
  • 1. NPK1:   Flowering plant (tobacco) homolog of Bck1

O-IV. Pak/Ste20 family
vertebrate:
  • 1. Pak: "p21-(Cdc42/Rac) activated kinase"
Saccharomyces cerevisiae:
    1. Ste20:   Product of gene required for pheromone response

O-V. NimA family
vertebrate:
    1. Nek1:   NimA-related kinase
  • 2. Nek2:   NimA-related kinase (Nlk1)
  • 3. Nek3:   NimA-related kinase
  • 4. Nrk2:   NimA-related kinase
  • 5. Stk1:   NimA-related kinase
Aspergillus nidulans:
    1. NIMA: Cell cycle control protein kinase
Drosophila melanogaster:
    1. Fused: Product of gene required for segment polarity

Table 1. *(continued).*

Trypanosoma brucei (Phylum Zoomastigina, Kingdom Protoctista):
    1. NrkA: Trypanosome protein kinase related to NimA
Saccharomyces cerevisaie:
    1. Kin3:   Putative protein kinase

O-VI. wee1/mik1 family
  vertebrate:
    1. Wee1Hu: Gene product able to complement S. pombe wee1 mutant
  Saccharomyces cerevisiae:
  * 1. Swe1:   Wee1 homolog from budding yeast
  Schizosaccharomyces pombe:
    1. SpWee1:   "Wee" size at division kinase; Cdc2 negative regulator
    2. Mik1:   "Mitosis inhibitory kinase", negative regulator of Cdc2

O-VII. Family of kinases involved in translational control
  vertebrate:
    1. HRI: "Heme-regulated eukaryotic initiation factor 2α kinase"
    2. PKR: "Double-stranded RNA-dependent kinase" (Tik)
  Saccharomyces cerevisiae:
    1. Gcn2:   Protein required for translational derepression

O-VIII. Raf family
  vertebrate:
    1. Raf-1:   Cellular homolog of retroviral oncogene product
    2. A-Raf: Oncogenic protein closely related to c-Raf
    3. B-Raf: Oncogenic protein closely related to c-Raf
  Drosophila melanogaster:
    1. DmRaf: Raf homolog
  Caenorhabditis elegans:
    1. CeRaf: Raf homolog; product of *lin-45* gene required for vulval differentiation
  Phylum Angiospermophyta (Kingdom Plantae):
    1. Ctr1:   Negative regulator of ethylene response pathway

O-IX. Activin/TGFβ receptor family
  A. Subfamily of type I receptors
    vertebrate:
      1. ActR-I: Type I receptor for activin and TGF-β (Tsk7L, SKR1, ALK-2)
    * 2. TSR-1:   Type I receptor for activin and TGFG-β (ALK-1)
    * 3. TGFβRI: Type I receptor TGF- (ALK-5)
    * 4. ActR-IB: Type I receptor for activin (ALK-4)
    * 5. BRK-1:   Type I receptor for BMP-2 and BMP-4 (ALK-3)
    * 6. ALK-6:   "Activin receptor-like kinase", type 6
    Drosophila melanogaster:
    * 1. DmAtr-I: Type I activin receptor homolog
    * 2. DmSax: Product of *saxophone* gene
  B. Subfamily of type II receptors
    vertebrate:
      1. ActRII: Type II receptor for activin
      2. ActRIIB: Type II receptor for activin
      3. TGFβRII: Type II receptor TGF-β
    * 4. C14:   Putative receptor kinase expressed in gonads
    Drosophila melanogaster:
    * 1. DmAtr-II: Type II activin receptor homolog
    Caenorhabditis elegans:
    * 1. DAF-4:   Larva development regulatory protein; BMP receptor
  C. Others
    Caenorhabditis elegans:
      1. DAF-1:   Product of gene required for vulval development

O-X. Flowering plant putative receptor kinase family
  Phylum Angiospermophyta (Kingdom Plantae):
    1. ZmPK1:   Putative receptor protein-serine kinase (maize)
    2. Srk: "S receptor kinase"; three distinct alleles: 2, 6, and 910 (Brassica)
    3. Tmk1:   Putative "Transmembrane receptor kinase" (Arabidopsis)
    4. Apk1:   Kinase that phosphorylates Tyr, Ser, and Thr (Arabidopsis)
  * 5. Nak: "Novel Arabidopsis Kinase" (Arabidopsis)
    6. Pro25:   Putative kinase selected for specificity to thylakoid membrane protein (Arabidopsis)
  * 7. Pto: Product of genen conferring pathogen resistance (tomato)
  * 8. Tmkl1:   Transmembrane protein with unusual kinase-like domain (Arabidopsis)
  * 9. Prk1:   Pollen-expressed receptor-like putative kinase (Petunia)

O-XI. Family of "mixed-lineage" kinases with leucine zipper domain
  vertebrate:
  * 1. Mlk1:   "Mixed lineage kinase", type 1
  * 2. Mlk2:   "Mixed lineage kinase", type 2
  * 3. Mlk3:   "Mixed lineage kinase", type 3 (PTK1, SPRK)

Table 1. *(continued).*

O-XII. Casein kinase I family

vertebrate:

	1. CK1α:	Casein kinase I, type alpha
	2. CK1β:	Casein kinase I, type beta
	3. CK1γ:	Casein kinase I, type gamma
	4. CK1δ:	Casein kinase I, type delta

Saccharomyces cerevisiae:

	1. Yck1:	Budding yeast casein kinase I homolog, type 1
	2. Yck2:	Budding yeast casein kinase I homolog, type 2
	3. Hrr25:	Kinase required for DNA repair

Schizosaccharomyces pombe:

*	1. Hhp1:	Fission yeast casein kinase I homolog, type 1
*	2. Hhp2:	Fission yeast casein kinase I homolog, type 2

O-XIII. PKN family of prokaryotic protein kinases

Myxococcus xanthus (Phylum Myxobacteria: Kingdom Prokaryotae):

	1. Pkn1:	Protein kinase homologous to eukaryotic kinases
	2. Pkn2:	Protein kinase required for maintenance of stationary phase cells and development

Other protein kinase family members (each with no known close relatives)

vertebrate:

	1. Mos:	Cellular homolog of retroviral oncogene product
	2. Pim1:	Proto-oncogene activated by murine leukemia virus
	3. Cot:	Product of oncogene expressed in human thyroid carcinoma
	4. Esk:	"Embryonal carcinoma STY kinase"; dual specificity (PTT)
*	5. GC kinase:	Kinase expressed in germinal center B cells
*	6. Slk:	STE20-related kinase
*	7. LIMK:	"LIM motif-containing kinase"
*	8. Tsk1:	"Testis-specific kinase"

Drosophila melanogaster:

	1. NinaC:	Product of gene essential for photoreceptor function
	2. Pelle:	Product of gene required for dorsalventral polarity
*	3. Nemo:	Product of gene required for rotation of photoreceptor clusters

Dictyostelium discoideum:

	1. SplA:	Spore lysis A protein kinase
	2. Dpyk2:	Developmentally-reguated tyrosine kinase, type 2

Ceratodon purpureus: (a moss)

	1. PhyCer:	Putative protein-tyrosine kinase encoded by a phytochrome gene

Saccharomyces cerevisiae:

	1. Cdc7:	"Cell-division-cycle" control gene product
	2. CDC15:	"Cell-division-cycle" control gene product
	3. Vps15:	Product of gene essential for sorting to lysosome-like vacuole
	4. Npr1:	Product of gene required for activity of ammonia-sensitive amino acid permeases
	5. Elm1:	Product of gene required for yeast-like cell morphology
	6. Ire1:	Required for Myo-inositol synthesis and signaling from ER to the nucleus
	7. Ykl516:	Putative protein kinase gene on chromosome XI
*	8. Ipl1:	Product of gene required for chromosome segregation

Schizosaccharomyces pombe:

	1. Ran1:	Product of gene required for normal meiotic function
	2. Chk1:	"Checkpoint Kinase" that links rad pathway to Cdc2
*	3. Csk1:	"Cyclin Suppressing Kinase"
*	4. RPK1:	"Regulatory cell proliferation kinase"

Entamoeba histolytica (Phylum Rhizopoda, Kingdom Protoctista):

	1. Ehmfk1:	Distant relative of Mos

Phylum Angiospermophyta (Kingdom Plantae):

	1. GmPK6:	Protein kinase homolog (soybean)
*	2. Tsl:	Product of Tousled gene required for normal leaf/flower development (Arabidopsis)

Yersinia psuedotuberculosis (Phylum Omnibacteria, Kingdom Prokaryotae):

	1. YpkA:	Enterobacterial protein kinase essential for virulence

known primary structures. The kinase domains are further divided into 12 smaller subdomains (indicated by Roman numerals), defined as regions never interrupted by large amino acid insertions and containing characteristic patterns of conserved residues (consensus line in Fig. 1).

Twelve kinase domain residues are recognized as being invariant or nearly invariant throughout the superfamily (conserved in over 95% of 370 sequences), and hence strongly implicated as playing essential roles in enzyme function. Using the type α cAMP-dependent protein kinase catalytic subunit (PKA-Ca) as a reference point, these are equivalent to Gly50 and Gly52 in subdomain I, Lys72 in subdomain II, Glu91 in subdomain III, Asp166 and Asn171 in subdomain VIB, Asp184 and Gly186 in subdomain VII, Glu208 in subdomain VIII, Asp220 and Gly225 in subdomain IX, and Arg280 in subdomain XI.

The patterns of amino acid residues found within subdomains VIB, VIII, and IX have been particularly well-conserved among the individual members of the dif-
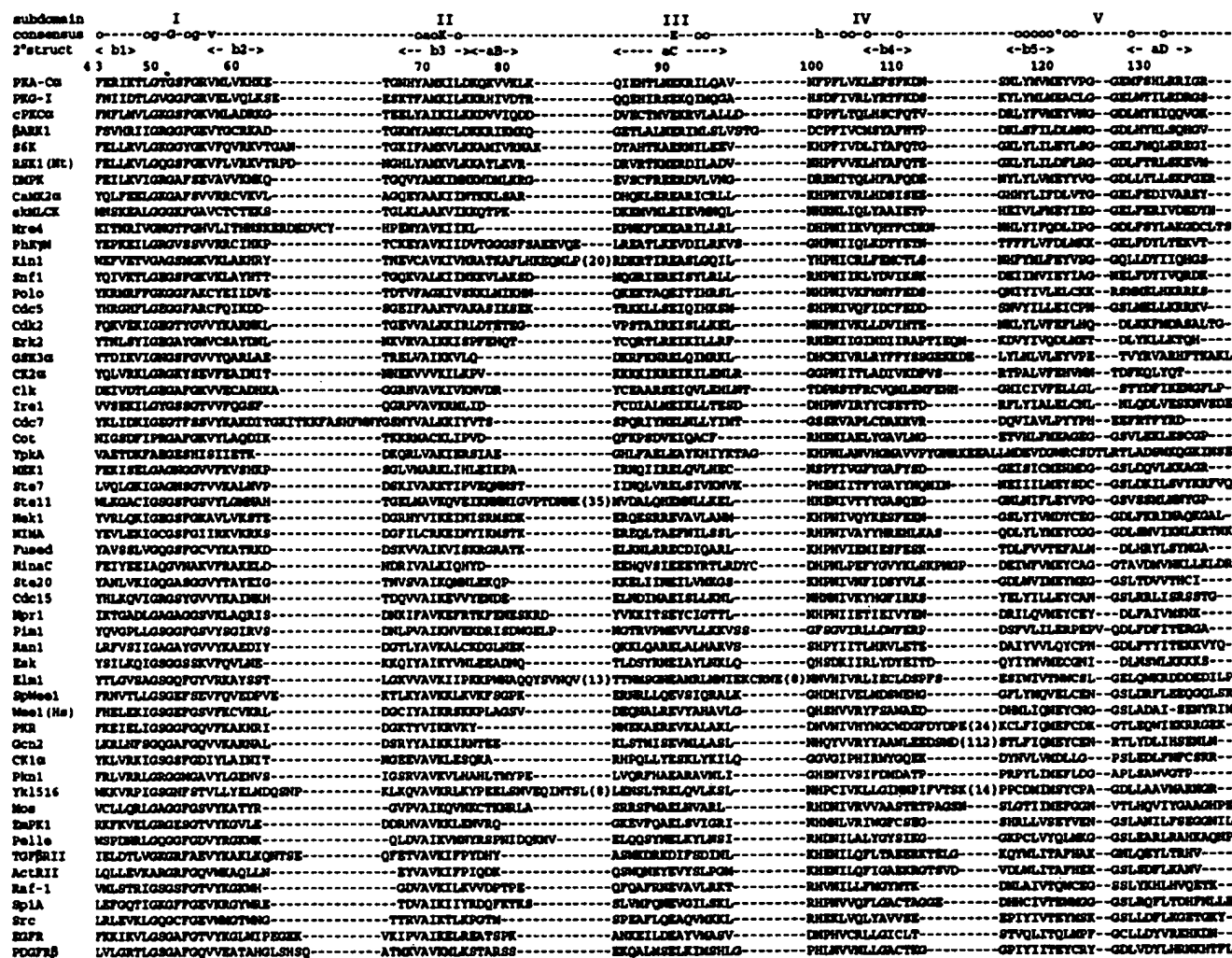
Figure 1. Multiple alignments of 60 kinase domains representative of members of the eukaryotic protein kinase superfamily. The abbreviated names used are as defined in Table 1. The single letter amino acid code is used and gaps are indicated by dashes. The entire sequences for the larger inserts are not shown, but excluded residues are indicated as numbers in brackets. Twelve distinct subdomains are indicated by Roman numerals. The consensus line is given according to the following code: uppercase letters, invariant residues, lowercase residues nearly invariant residues; o, positions conserving nonpolar residues; *, positions conserving polar residues; +, positions conserving small residues with near neutral polarity. Residues corresponding to the numbered β-strands (b) and α-helices (a) in PKA-Cα are indicated in the 2· structure line.

ferent protein kinase families and these motifs have been targeted most frequently in PCR-based homology cloning strategies aimed at identifying new family members.

Relationship between conserved subdomains, higher order structure, and catalytic mechanism

The homologous nature of the kinase domains implies that they all fold into topologically similar 3-dimensional core structures and impart phosphotransfer according to a common mechanism. The larger inserts found within some kinase domains are likely to represent surface elements that do not disrupt the basic core structure. With the solution of the crystal structure of mouse PKA-Cα, in a binary complex with a pseudosubstrate peptide inhibitor (PKI 5-24; TTYADFIASGRTGRRNAIHD, the underlined Ala substituting for the Ser phosphoacceptor), the general topology of a protein kinase catalytic core struc-

ture was revealed for the first time (25, 26). Later, structures of ternary complexes of PKA-Cα, the pseudosubstrate inhibitor, and either MgATP or MnAMP-PNP (an MgATP analog) were solved (27, 28). As a consequence of these studies, precise functional roles for most of the highly conserved kinase domain residues have now been assigned.

The kinase domain of PKA-Cα folds into a two-lobed structure (**Fig. 2**). The smaller, NH2- terminal lobe, which includes subdomains I-IV, is primarily involved in anchoring and orienting the nucleotide. This lobe has a predominantly antiparallel β-sheet structure that is unique among nucleotide binding proteins. The larger COOH-terminal lobe, which includes subdomains VIA-XI, is largely responsible for binding the peptide substrate and initiating phosphotransfer. It is predominantly α-helical in content. Subdomain V residues span

subdomain	VIA	VIB	VII	VIII
consensus	--o-------o--*o--+o-ooh--	-oohrDok+-Nooo-	-oko+Dfgo+-	-g+--o-+pHoo-
2°struct	<------- αH ------->	<b6> < b7>	< b8 > <b9>	
	140 150 160	170	180 190	200 210

(sequence alignment rows — Figure 1 continued)

Figure 1 (contd.).

the two lobes. The deep cleft between the two lobes is recognized as the site of catalysis. The crystal structures of four additional eukaryotic protein kinase superfamily members—cyclin-dependent kinase 2 (Cdk2) (29), p42 MAP kinase (Erk2) (30), twitchin kinase (31), and casein kinase I (32)—have been reported more recently, and as expected, their kinase domains were found to fold into two-lobed structures topologically very similar to the catalytic core of PKA-Cα. Notable differences, however, were found in the regions corresponding to subdomain VIII in the Cdk2 and Erk2 structures, apparently reflecting the fact that these are structures of enzymes in an inactive state (see below). The twitchin structure is also of an inactive enzyme, but in this case it is inactive due to the presence of an autoinhibitory peptide sequence, which lies on the COOH-terminal side of the kinase domain and folds back into the active site cleft between the two lobes (31). This peptide apparently forces the two

lobes to rotate almost 30° with respect to one another, and in this configuration inactive twitchin is more similar to the open configuration of PKA-Cα without PKI (33). In both twitchin and Cdk2 the α-helix C in subdomain III also adopts a different position to that of helix C in PKA-Cα. Unfortunately, no structure of a protein-tyrosine kinase catalytic domain was available at the time of writing (see "Note added in proof"), but the ease with which it has been possible to model the kinase domain of the EGF receptor protein-tyrosine kinase on to that of the PKA-Cα emphasizes that the structure of the protein-tyrosine kinases will be similar to that of the protein-serine kinases (34)

The conserved kinase subdomains correspond quite well to precise units of higher order structure. The functions of the individual subdomains will be discussed briefly later on a subdomain-by-subdomain basis, making reference to the crystal structure of PKA-Cα and

Figure 1 (contd.).

drawing attention to the proposed roles of the nearly invariant amino acid residues (25-27, 28) and other residues of interest. For more detailed information, the reader is referred to recent reviews on the structure of PKA-Cα (35-37) and to an excellent comparative review of the structures of PKA-Cα, Erk2, and Cdk2 (38).

Subdomain I, at the NH2 terminus of the kinase domain, contains the consensus motif Gly-x-Gly-x-x-Gly-x-Val (starting with Gly50 in PKA-Cα). The kinase domain NH2-terminal boundary occurs seven positions upstream of the first glycine in the consensus, where a hydrophobic residue is usually found. Subdomain I residues fold into a β-strand-turn-β-strand structure encompassing β-strands 1 and 2, and this structure acts as a flexible flap or clamp that covers and anchors the non-transferable phosphates of ATP. The backbone amides of Ser53, Phe54, and Gly55 form hydrogen bonds with ATP β- phosphate oxygens. Leu49 and Val57 contribute to a hydrophobic pocket that encloses the adenine ring of ATP.

Subdomain II contains the invariant Lys (Lys72 in PKA-Cα), which has long been recognized as being essential for maximal enzyme activity. This Lys lies within β-strand 3 of the small lobe, and helps anchor and orient ATP by interacting with the α- and β- phosphates. In addition, Lys72 forms a salt bridge with the carboxyl group of the nearly invariant Glu91 in subdomain III. Ala70 contributes to the hydrophobic adenine ring pocket. In PKA-Cα, β-strand 3 is followed immediately by α-helix B, which, judging from the sequence alignment, appears to be quite a variable structure among the protein kinases. Indeed, this α- helix is absent in the Cdk2 and Erk2 crystal structures.

Subdomain III represents the large α- helix C in the small lobe. The nearly invariant Glu residue (Glu91 in PKA-Cα) is centrally located in this helix and helps stabilize the interactions between Lys72 and the α- and β-phosphates of ATP. Subdomain IV corresponds to the hydrophobic β-strand 4 in the small lobe. This subdomain contains no invariant or nearly invariant residues
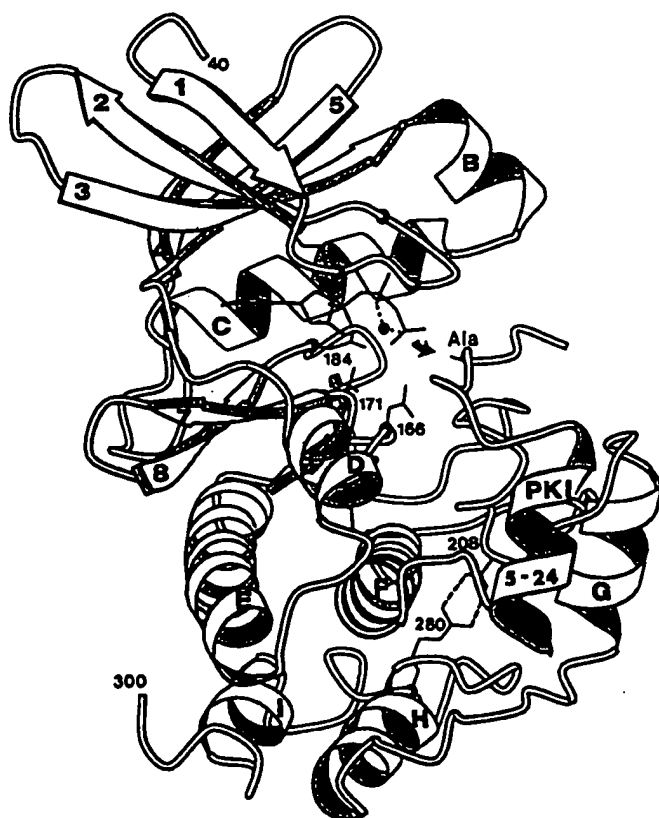
Figure 2. Ribbon diagram of the catalytic core of PKAα (residues 40–300) in a ternary complex with MgATP and pseudosubstrate peptide inhibitor (PKI -5-24). Invariant or nearly-invariant residues (Gly50, Gly52, Gly55, Lys72, Glu91, Asp166, Asn171, Asp184, Glu208, Asp220, and Arg280) are indicated by dots along the ribbon diagram. Side chains are shown for Lys72, Asp166, Asn171, Asp184, Glu208, and Arg280. β-strands and α-helices are indicated by flat arrow and helices, respectively, and are numbered according to Knighton et al. (26). The small arrow indicates the site of phosphotransfer with the Ala in PKI substituting for the phosphoacceptor Ser in the true substrate. (Reproduced, with permission, from Taylor et al. (36)).

and does not appear to be directly involved in catalysis or substrate recognition.

Subdomain V links the small and large lobes of the catalytic subunit and consists of the very hydrophobic β-strand 5 in the small lobe, the small α-helix D in the large lobe, and an extended chain that connects them. Three residues in the connecting chain of PKA-Cα, Glu121, Val123, and Glu127 help anchor ATP by forming hydrogen bonds with either the adenine or the ribose ring. Met120, Tyr122, and Val123 contribute to the hydrophobic pocket surrounding the adenine ring. Glu127 also participates in peptide binding by forming an ion pair with an Arg in the pseudosubstrate site of the PKA inhibitor peptide. This represents the first Arg in the PKA substrate recognition consensus Arg-Arg-x-Ser*-Hydrophobic.

Subdomain VIA folds into the large hydrophobic α-helix E that extends through the large lobe. None of the

residues in helix E appear to interact directly with either MgATP or peptide substrate; hence this part of the molecule appears to act mainly as a support structure. Subdomain VIB folds into the small hydrophobic β-strands 6 and 7 with an intervening loop. Included here are two invariant residues (Asp166 and Asn171 in PKA-Cα) that lie within the consensus motif His-Arg-Asp-Leu-Lys-x-x-Asn (HRDLKxxN). The loop has been termed the catalytic loop because Asp166 within the loop has emerged as the likely candidate for the catalytic base, accepting the proton from the attacking substrate hydroxyl group during an in- line phosphotransfer mechanism. Lys168 in the loop (substituted by Arg in the conventional protein-tyrosine kinases) may help facilitate phosphotransfer by neutralizing the negative charge of the γ-phosphate during transfer. The side chain of Asn171 helps to stabilize the catalytic loop through hydrogen bonding to the backbone carbonyl of Asp166 and also acts to chelate the secondary Mg^{2+} ion that bridges the α- and γ-phosphates of the ATP. The carbonyl group of Glu170 forms a hydrogen bond with an ATP ribose hydroxyl group. Glu170 also participates in substrate binding by forming an ion pair with the second arginine of the peptide recognition consensus.

Subdomain VII folds into a β-strand-loop-b-strand structure, encompassing β-strands 8 and 9. The highly conserved DFG triplet, corresponding to Asp184-Phe185-Gly186 in PKA-Cα, lies in the loop that is stabilized by a hydrogen bond between Asp184 and Gly186. Asp184 chelates the primary activating Mg^{2+} ions that bridge the β- and γ-phosphates of the ATP, and thereby helps to orient the γ-phosphate for transfer. In Cdk2, β-strand 9 is replaced with a small α-helix designated αL12. However, it is unclear whether this helical character is maintained when Cdk2 is in its active conformation.

Subdomain VIII, which includes the highly conserved Ala-Pro-Glu ('APE') motif (residues 206-208 in PKA-Cα), folds into a tortuous chain that faces the cleft. Residues lying 7-10 positions immediately upstream of the APE motif are characteristically well-conserved among the members of different protein kinase families. The nearly invariant Glu corresponding to PKA-Cα Glu208 forms an ion pair with an invariant Arg (Arg280 in PKA-Cα) in subdomain XI, thereby helping to stabilize the large lobe.

Subdomain VIII appears to play a major role in recognition of peptide substrates. Several PKA-Cα subdomain VIII residues participate in binding the pseudosubstrate inhibitor peptide. Leu198, Cys199, Pro202, and Leu205 of PKA-Cα provide a hydrophobic pocket that accommodates the side chain of the hydrophobic residue at position +1 of the substrate consensus (Ile for the inhibitor peptide). Gly200 forms a hydrogen bond with the same Ile residue. Glu203 forms two ion pairs with the Arg in the high-affinity binding region of the inhibitor peptide.

Many protein kinases are known to be activated by phosphorylation of residues in subdomain VIII. In PKA-Cα, maximal kinase activity requires phosphorylation of Thr197, probably occurring through an intermolecular autophosphorylation mechanism (39). In the crystal structure, phosphate oxygens of phospho-Thr197 form hydrogen bonds with the charged side chains of Arg165, Lys189, and the hydroxyl group of Thr195, and thereby may act to stabilize the subdomain VIII loop in an active conformation permitting proper orientation of the substrate peptide. For members of the Erk (MAP) kinase family, phosphorylation of both a Thr and a Tyr

residue in subdomain VIII (mediated by members of the MEK kinase family) is required for activation. In the crystal structure determined for Erk2, these residues (Thr183 and Tyr185) were not phosphorylated and thus the enzyme was in an inactive state (unlike the PKA-Cα structure). The unphosphorylated Tyr185 is buried in a hydrophobic pocket, and interactions with Tyr185 are apparently required to hold the enzyme in the inactive state. Mutation of Tyr185, however, does not activate the enzyme, and so phosphorylation of Tyr185 must also play a role in activation. Unphosphorylated Erk2 appears to be inactive because residues required for catalysis are not properly oriented, and because its conformation results in a partial steric block to substrate binding. During activation of Erk2, Tyr185 phosphorylation precedes Thr183 phosphorylation; therefore, binding of MEK to Erk2 may alter the conformation of the subdomain VIII loop, thereby exposing Tyr185 for phosphorylation by MEK. Interaction of phospho-Tyr185 with surface residues would then allow the subdomain VIII loop to adopt the active conformation (30). Subsequent phosphorylation of the exposed Thr183 may activate the enzyme fully by promoting correct alignment of the catalytic residues. From the crystal structure of Cdk2, likewise in an inactive unphosphorylated state, the subdomain VIII loop appears to be in a conformation that would inhibit enzyme activity by sterically blocking the presumed protein substrate binding cleft (29). Phosphorylation of Thr160 in the Cdk2 subdomain VIII, mediated by MO15 (CAK), presumably would act to remove this inhibition by stabilizing the loop in an active conformation similar to that found in PKA-Cα. Cyclin binding to the NH2-terminal lobe is also needed to activate Cdk2, and this may cause rotation of the NH2-terminal domain resulting in correct alignment of catalytic residues.

Subdomain IX corresponds to the large α- helix F of the large lobe. The nearly invariant Asp corresponding to PKA-Cα Asp220 lies in the NH2-terminal region of this helix and acts to stabilize the catalytic loop by hydrogen bonding to the backbone amides of Arg165 and Tyr164 that precede the loop. Glu230 of PKA-Cα forms an ion pair with the second Arg of the peptide recognition consensus. PKA-Cα residues 235-239 are all involved in hydrophobic interactions with the inhibitor peptide.

Subdomain X is the most poorly conserved subdomain and its function is obscure. In the crystal structure of PKA-Cα, it corresponds to the small α-helix G that occupies the base of the large lobe. Members of the Cdk, Erk (MAP), GSK3, and Clk kinase families (the C-M-G-C group) all have rather large insertions between subdomains X and XI, whose functional significance is presently unclear. Subdomain XI extends to the COOH-terminal end of the kinase domain. The most notable feature here is the nearly invariant Arg corresponding to Arg280 in PKA-Cα, which lies between α-helices H and I. The COOH-terminal boundary of the kinase domain is still poorly defined. For many protein-serine kinases, the consensus motif His-x-Aromatic-Hydrophobic is found beginning 9-13 residues downstream of the invariant Arg. For protein-tyrosine kinases, a hydrophobic amino acid lying 10 positions downstream of the invariant Arg appears to define the COOH-terminal boundary.

The amphipathic α-helix A of PKA-Ca (residues 15-35; not shown in Fig. 2), though lying outside of the conserved catalytic core on the NH2-terminal side, appears to be an important feature found in many protein

kinases (40). This helix spans the surface of both lobes of the core structure and complements and stabilizes the hydrophobic cleft between the two lobes. The A-helix motif appears to be present in many other protein kinases including members of the protein kinase C family and the Src family of protein-tyrosine kinases (40).

CLASSIFICATION OF EUKARYOTIC PROTEIN KINASES

To facilitate analysis and management of this large superfamily we have devised the classification scheme shown in Table 1, which subdivides the known members of the eukaryotic protein kinase superfamily into distinct families that share basic structural and functional properties. Phylogenetic trees derived from an alignment of kinase domain amino acid sequences (essentially an expanded version of Fig. 1) served as the basis for this classification. Thus, the sole consideration was similarity in kinase domain amino acid sequence. When considered alone, however, this property has been a good indicator of other characteristics held in common by the different members of the family.

Protein kinases whose entire kinase domain amino acid sequence had been published by July 1993 were included in phylogenetic analysis (as well as a few others made available at that time through sequence databases). If a given kinase domain sequence had been determined from more than one species among the vertebrates (i.e., orthologous gene products), only one representative (usually human) was included in the analysis. This policy was not used for the other phyla, however, because of greater divergences between the species and, hence, the sequences. The kinase domain phylogenies were inferred using the principle of maximum parsimony according to the PAUP software package developed by Swofford (41). Minimum-length trees were found using PAUP's 'heuristic' search method with branch swapping by the 'tree bisection-reconnection' strategy. Equal weights were given for all amino acid substitutions. Because multiple minimum-length trees were found, a consensus tree was calculated according to the method of Adams (cited in ref 41) in order to show branching ambiguities.

To accommodate the large numbers of sequences, it was necessary to construct five separate trees. Initially, a skeleton tree of 99 kinases was obtained (Fig. 3A). The skeleton tree included only representative members from each of four large groups of protein kinases, each consisting of multiple related families known from previous work to cluster together in the tree. These four groups are designated: 1) the AGC group, which includes the cyclic-nucleotide-dependent family (PKA and PKG), the protein kinase C (PKC) family, the β-adrenergic receptor kinase (βARK) family, the ribosomal S6 kinase family, and other close relatives; 2) the CaMK group, which includes the family of protein kinases regulated by calcium/calmodulin, the Snf1/AMPK family, and other close relatives; 3) the CMGC group, which includes the family of cyclin-dependent kinases, the Erk (MAP) kinase family, the glycogen synthase 3 (GSK3) family, the casein kinase II family, the Clk (Cdk-like kinase) family, and other close relatives; and 4) the 'conventional' protein-tyrosine kinase (PTK) group. Separate trees (Fig. 3B-E) were later obtained for each of the four large kinase groups, and contain all members of the groups whose sequences were available at the time of analysis.
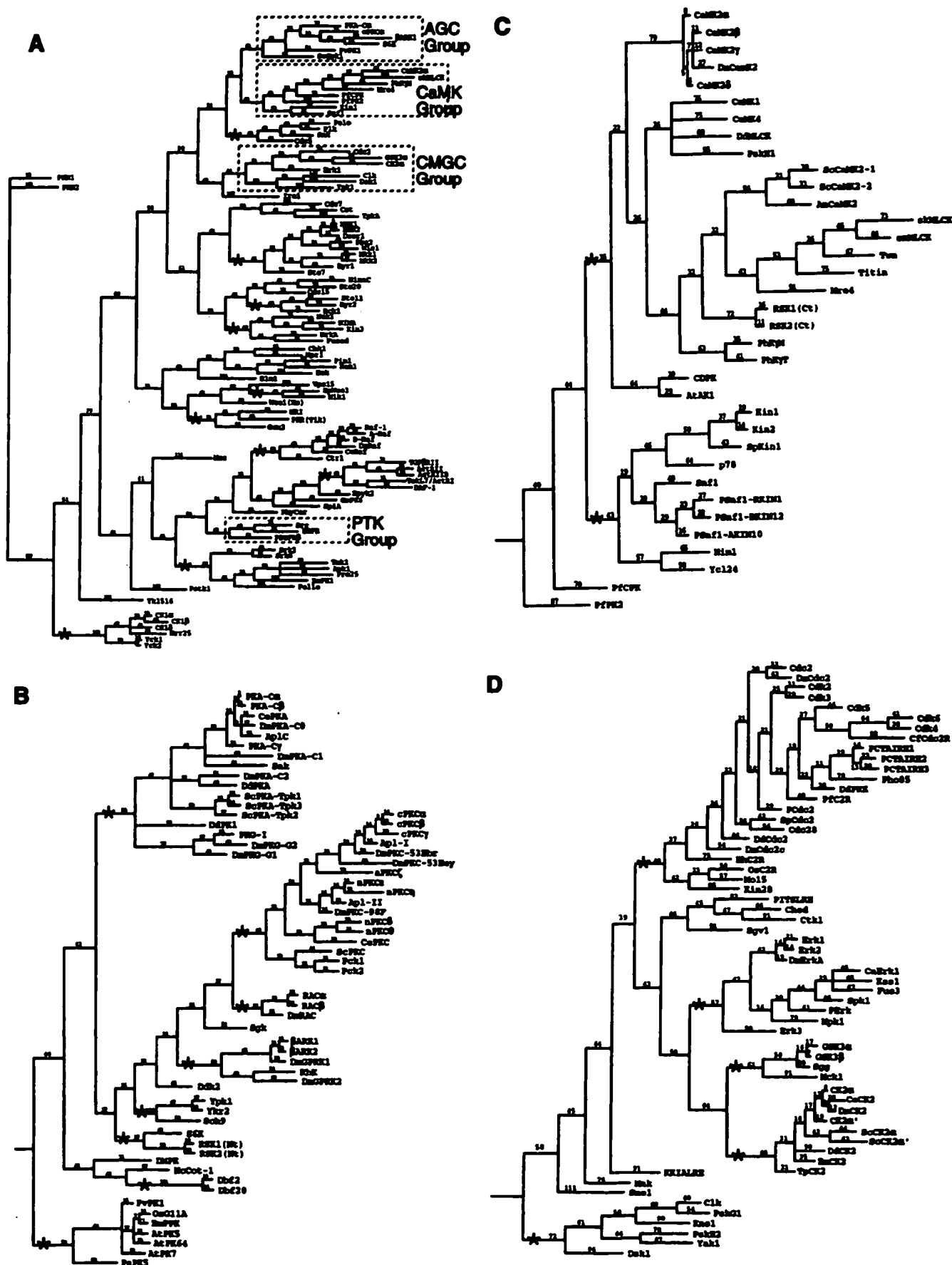
E
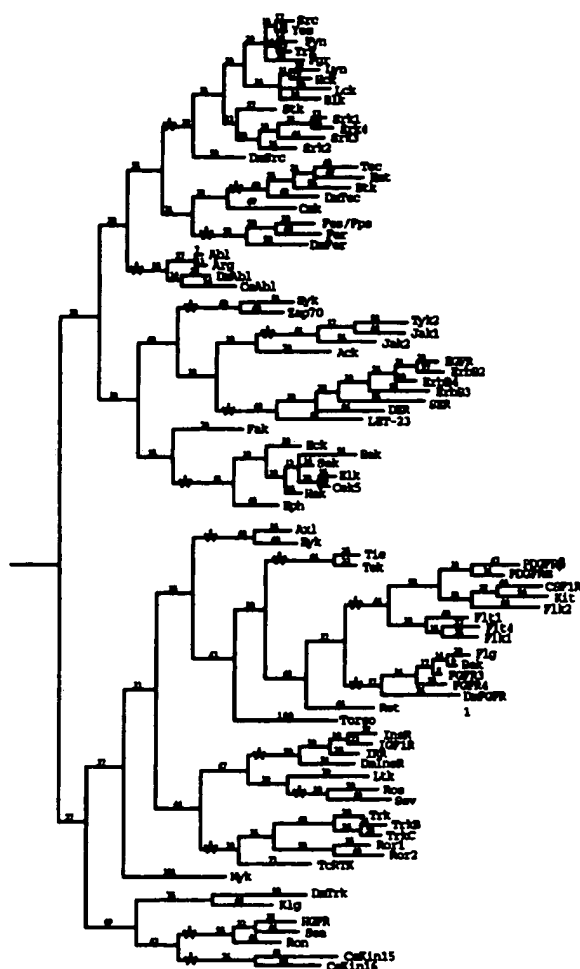
Figure 3. Phylogenetic trees of the eukaryotic protein kinase superfamily inferred from kinase domain amino acid sequence alignments. The abbreviated nomenclature is the same used in Table 1. *A)* 'Skeleton' tree showing 99 protein kinases. Positions of 4 clusters (AGC, CaMK, CMGC, and PTK) containing protein kinases representative of larger groups are indicated in the skeleton tree. *B)* AGC group tree of 59 protein kinases including PKA, PKG, and PKC and other close relatives. *C)* CaMK group tree of 35 protein kinases including the calcium/calmodulin-regulated enzymes. *D)* CMGC group tree of 59 protein kinases including the cyclin-dependent kinases. *E)* PTK group tree of 90 conventional protein-tyrosine kinases. Tree A is unrooted and drawn with Pkn1 and Pkn2 as outgroups. Outgroups of two or more distantly related protein kinases (not shown) were included in the analysis of trees B-E to provide a rooting point. Asterisks (*) in all trees indicate branches leading to defined protein kinase families listed in Table 1. Branch lengths indicate number of amino acid substitutions required to reach hypothetical common ancestors at internal nodes.

It can be reasonably surmised that the protein kinases having closely related catalytic domains, and thus defining a family, represent products of genes that have undergone relatively recent evolutionary separations. Given this, it should come as no surprise that members of a given family tend also to share related functions. This is manifest by similarities in overall structural topology, mode of regulation, and substrate specificity. The details of the common properties exhibited by the members of the various kinase families can best be gleaned from studying the information outlined in the individual entries section of the *Protein Kinase Factsbook* (42). Some of the most salient relationships are discussed below.

The AGC group protein kinases tend to be basic amino acid-directed enzymes, phosphorylating substrates at Ser/Thr residues lying very near Arg and Lys. For the cyclic nucleotide-dependent and ribosomal S6 kinase families, the preferred substrates have basic residues lying in specific positions NH2-terminal to the phosphate acceptor. Preferred substrates for the PKC and RAC families have basic residues on both the NH2- and COOH-terminal sides of the acceptor (43). The G-protein-coupled receptor kinases (βARK and RhK) appear to break this rule, however, as they are reported to prefer synthetic peptide substrate residues located within an acidic environment. Little substrate information is available for the other families in this group.

The CaMK group protein kinases also tend to be basic amino acid- directed, and in this regard it is notable that the AGC and CaMK groups fall near one another in the phylogenetic tree. CaMK1, CaMK2, CaMK4, MLCK, CDPK, and AMPK are all reported to prefer substrates with basic residues at specific positions NH2-terminal to the acceptor site, whereas EF2K and PhK prefer sites with basic residues at both NH2- and COOH-terminal locations. Many, but not all, of the CaMK group protein kinases are known to be activated by Ca²⁺/calmodulin binding to a small domain located just COOH-terminal to the catalytic domain, e.g., CaMK1, CaMK2, CaMK4, PhKγ, MLCK, and twitchin. These enzymes and their close relatives are grouped together in a large family within the CaMK group. Also included in this family are a subfamily of plant enzymes (represented by CDPK) that contain an intrinsic calmodulin-like domain that confers Ca²⁺-dependent activation. The other family within the CaMK group is the Snf1/AMPK family. Within this family, substrate specificity determinant information has been obtained only for the AMP-activated protein kinase, which also shows a requirement for an NH2-terminal basic residue. The other major category of protein-serine kinases is the CMGC group. For the most part, these are proline-directed enzymes, phosphorylating substrates at sites lying in Pro-rich environments. Available data for Cdc2 and Cdk2 indicate that members of the cyclin-de-

pendent kinase family require phosphate acceptors lying immediately NH₂-terminal to a Pro. A similar requirement is indicated for the Erk (MAP) kinase family. The situation for the GSK3 family is more complicated, but most known acceptor sites lie within Pro-rich regions. The structures of Cdk2 and Erk2 indicate that the pocket for the +1 residue is shallower than in PKA-Cα due to the replacement of Leu205 by an Arg, which is bulkier and precludes binding of the larger hydrophobic amino acids. In addition, the unique secondary amide group of Pro may make special interactions (44). The casein-kinase II family enzymes fail to conform to the proline-directed specificity exhibited by the other major families of this group, showing instead a strong preference for Ser residues located NH₂-terminal to a cluster of acidic residues. The CMGC group protein kinases have larger-than-average kinase domains due to insertions between subdomains X and XI, whose functional significance is unknown.

The conventional protein-trosine kinase group includes a large number of enzymes with quite closely related kinase domains that specifically phosphorylate on Tyr residues (i.e., they cannot phosphorylate Ser or Thr). These enzymes, first recognized among retroviral oncoproteins, have been found only in metazoan cells where they are widely recognized for their roles in transducing growth and differentiation signals. Included in this group are more than a dozen distinct receptor families made up of membrane-spanning molecules that share similar overall structural topologies, and nine nonreceptor families also composed of structurally similar molecules. The specificity determinants surrounding the Tyr phosphoacceptor sites have yet to be firmly established for these enzymes, but Glu residues either on the NH₂- or COOH-terminal side of the acceptor are often preferred. This group is labeled "conventional" to distinguish it from other protein kinases (including Spk1, Clk, the MEK/Ste7 family members, Wee1/Mik1, ActRII, Hrr25, Esk, and Sp1A/DPyk2) reported to exhibit a dual specificity, that is, being capable of phosphorylating both Tyr and Ser/Thr residues (45). However, in most cases dual specificity has been observed only for autophosphorylation reactions in vitro, and the only dual specificity protein kinases that are known to be able to phosphorylate a substrate on Ser/Thr and Tyr are members of the MEK family. Considered as a group, these dual-specificity protein kinases are not particularly closely related to the conventional PTKs. Indeed, they seem to map throughout the phylogenetic tree (45), suggesting that the ability to autophosphorylate on Tyr may have had many independent origins during the evolutionary history of the superfamily.

The protein kinases falling outside the four major groups are a mixed bag. Although the individual members within the defined families found in this "other" category clearly are related to one another through both structure and function, it is difficult to make broader generalizations that could group any of these families together into a larger category. As far as substrate specificity determinants go, little is known about most "other" category protein kinases, due primarily to their rather recent discovery and the paucity of known physiological substrates. The casein kinase I family members, however, have been shown to prefer Ser/Thr residues located COOH-terminal to a phosphoserine or phosphothreonine, although a stretch of acidic residues may substitute.

Also, the family of protein kinases involved in translational control (HRI, PKR/Tik, Gcn2) appear to be basic amino acid-directed enzymes preferring Ser residues lying NH₂- terminal to an Arg. Finally, as mentioned previously, the MEK/Ste7 family protein kinases and Wee1/Mik1 protein kinases exhibit a dual specificity.

Although this classification is based solely on catalytic domain sequences, members of families defined by this means are usually closely related in regions lying outside the catalytic domains and in many cases have been shown to possess very similar functions. Thus, intercalation of newly discovered protein kinases into this classification should allow one to make useful predictions about the functions of such enzymes.

FUTURE PROSPECTS

The rate of protein kinase discovery still shows no signs of abating. In addition to the continuing successes of homology-based approaches, genomic sequencing projects are beginning to make significant contributions. For instance, the sequences of two entire budding yeast chromosomes (46, 47) and a ~2 Mb stretch of C. elegans chromosome III (48) have revealed a number of new putative protein kinase genes. As genome sequencing projects gather speed, the number of new protein kinase genes discovered in this way will undoubtedly mushroom. This explosion of sequence data is making it increasingly difficult to manage protein kinase databases of the sort described here. Programs designed to align and derive relatedness trees are currently unable to handle the large number of available kinase domain sequences. New data handling programs will have to be developed to cope with large numbers of sequences like those of the eukaryotic protein kinase superfamily.

Protein kinase catalytic domain structures will continue to be solved. The first structure of a conventional protein-tyrosine kinase will be available shortly (see "Note added in proof"), and this should reveal how Tyr is selected as an acceptor amino acid vs. Ser/Thr. Such structures will enable comparative analysis to be carried out at the 3-dimensional level, and allow predictions of structures from primary sequences. Structural comparisons of catalytic domains with bound peptide substrates will also provide insights into substrate specificity. Most protein kinases show some degree of primary sequence specificity, and new methods are being developed to determine consensus sequence specificities for individual protein kinases (44). With such consensus information the structural basis for the binding of a preferred peptide sequence to the cognate substrate binding site can then be deduced. In the future, it may be possible to model the 3-dimensional structure of a novel protein kinase catalytic domain with sufficient accuracy to be able to deduce the preferred primary sequence surrounding the hydroxyamino acid it phosphorylates, which in turn will allow one to predict what proteins might be its substrates from the increasingly complete database of protein sequences. 𝔽𝕁

Note added in proof: The crystal structure of the tyrosine kinase domain of the insulin receptor has now appeared (Hubbard, S. R., Wei, L., Ellis, L., and Hendrickson, W. A. (1994) *Nature* 372, 746–754).

REFERENCES

1. Hanks, S. K., Quinn, A. M., and Hunter, T. (1988) The protein kinase family: conserved features and deduced phylogeny of the catalytic domains. *Science* **241**, 42–52
2. Hanks, S. K. (1991) Eukaryotic protein kinases. *Curr. Opin. Struct. Biol.* **1**, 369–3833.
3. Hanks, S. K., and Quinn, A. M. (1991) Protein kinase catalytic domain sequence database: identification of conserved features of primary structure and classification of family members. *Methods Enzymol.* **200**, 38–62
4. Hunter, T. (1987) A thousand and one protein kinases. *Cell* **50**, 823–8295.
5. Hunter, T. (1994) 1001 protein kinases redux: towards 2000. *Seminars Cell Biol.* In press
6. Muñoz-Dorado, J., Inouye, S., and Inouye, M. (1991) A gene encoding a protein serine/threonine kinase is required for normal development of *M. xanthus*, a gram-negative bacterium. *Cell* **67**, 995–1006
7. Galyov, E. E., Hakansson, S., Forsberg, A., and Wolf-Watz, H. (1993) A secreted protein kinase of *Yersinia pseudotuberculosis* is an indispensable virulence determinant. *Nature* **361**, 730–732
8. Hoekstra, M. F., DeMaggio, A. J., and Dhillon, N. (1991) Genetically identified protein kinases in yeast. Part 1: transcription, translation, transport and mating. *Trends Genet.* **7**, 256–261
9. Alex, L. A., Simon, M. J. (1994) Protein histidine kinases and signal transduction in prokaryotes and eukaryotes. *Trends Genet.* **10**, 133–136
10. Chang, C., Kwok, S. F., Bleecker, A. B., and Meyerowitz, E. M. (1993) *Arabidopsis* ethylene-response gene *ETR1*: similarity of product to two-component regulators. *Science* **262**, 539–544
11. Ota, I. M., and Varshavsky, A. (1993) A yeast protein similar to bacterial two-component regulators. *Science* **262**, 566–569
12. Maeda, T., Wurgler-Murphy, S. M., and Saito, H. (1994) A two-component system that regulates an osmosensing MAP kinase cascade in yeast. *Nature* **369**, 242–245
13. Popov, K. M., Zhao, Y., Shimomura, Y., Kuntz, M. J., and Harris, R. A. (1992) Branched-chain α-ketoacid dehydrogenase kinase. *J. Biol. Chem.* **267**, 13127–13130
14. Popov, K. M., Kedishvili, N. Y., Zhao, Y., Shimomura, Y., Crabb, D. W., and Harris, R. A. (1993) Primary structure of pyruvate dehydrogenase kinase establishes a new family of eukaryotic protein kinases. *J. Biol. Chem.* **268**, 26602–26606
15. Maru, Y., and Witte, O. N. (1991) The BCR gene encodes a novel serine/threonine kinase activity within a single exon. *Cell* **67**, 459–468
16. Beeler, J. F., LaRochelle, W. J., Chedid, M., Tronick, S. R., and Aaronson, S. A. (1994) Prokaryotic expression cloning of a novel human tyrosine kinase. *Mol. Cell. Biol.* **14**, 982–988
17. Huang, J. M., Wei, Y. F., Kim, Y. H., Osterberg, L., and Matthews, H. R. (1991) Purification of a protein histidine kinase from the yeast *Saccharomyces cerevisiae*. The first member of this class of protein kinases. *J. Biol. Chem.* **266** 9023–9031
18. Stock, J. B., Ninfa, A. J., and Stock, A. M. (1989) Protein phosphorylation and regulation of adaptive responses in bacteria. *Microbiol. Rev.* **53**, 450–490
19. Cozzone, A. J. (1993) ATP-dependent protein kinases in bacteria. *J. Cell. Biochem.* **51**, 7–13
20. Saier, M. H. (1993) Introduction: protein phosphorylation and signal transduction in bacteria. *J. Cell. Biochem.* **51**, 1–6
21. Reizer, J., Romano, A. H., and Deutscher, J. (1993) The role of phosphorylation of HPr, a phosphocarrier protein of the phosphotransfer system, in the regulation of carbon metabolism in gram-positive bacteria. *J. Cell. Biochem.* **4751**, 19–24
22. LaPorte, D. C. (1993) Isocitrate dehydrogenase phosphorylation cycle: regulation and enzymology. *J. Cell. Biochem.* **51**, 14–18
23. Muñoz-Dorado, J., Inouye, S., and Inouye, M. (1993) Eukaryotic-like protein serine/threonine kinases in *Myxococcus xanthus*, a developmental bacterium exhibiting social behavior. *J. Cell. Biochem.* **51**, 29–33
24. Zhang, C.-C. (1993) A gene encoding a protein-related to eukaryotic protein kinases from the filamentous heterocystous cyanobacterium *Anabaena* PCC7120. *Proc. Natl. Acad. Sci. USA* **90**, 11840–11844
25. Knighton, D. R., Zheng, J., Ten Eyck, L. F., Xuong, N.-H., Taylor, S. S., and Sowadski, J. M. (1991) Structure of a peptide inhibitor bound to the catalytic subunit of cyclic adenosine monophosphate-dependent protein kinase. *Science* **253**, 414–420
26. Knighton, D. R., Zheng, J., Ten Eyck, L. F., Ashford, V. A., Xuong,

N.-H., Taylor, S. S., and Sowadski, J. M. (1991) Crystal structure of the catalytic subunit of cyclic adenosine monophosphate-dependent protein kinase. *Science* **253**, 407–420
27. Bossemeyer, D., Engh, R. A., Kinzel, V., Ponstingl, H., and Huber, R. (1993) Phosphotransferase and substrate binding mechanism of the cAMP-dependent protein kinase catalytic subunit from porcine heart as deduced from the 2.0 Å structure of the complex with Mn²⁺ adenyl imidodiphosphate and inhibitor peptide PKI(5-24). *EMBO J.* **12**, 849–859
28. Zheng, J., Knighton, D. R., ten Eyck, L. F., Karlsson, R., Xuong, N., Taylor, S. S., and Sowadski, J. M. (1993) Crystal structure of the catalytic subunit of cAMP-dependent protein kinase complexed with MgATP and peptide inhibitor. *Biochemistry* **32**, 2154–2161
29. De Bondt, H. L., Rosenblatt, J., Jancarik, J., Jones, H. D., Morgan, D. O., and Kim, S. (1993) Crystal structure of cyclin-dependent kinase 2. *Nature* **363**, 595–602
30. Zhang, F., Strand, A., Robbins, D., Cobb, M. H., and Goldsmith, E. J. (1994) Atomic structure of the MAP kinase ERK2 at 2.3 Å resolution. *Nature* **367**, 704–711
31. Hu, S.-H., Parker, M. W., Lei, J. Y., Wilce, M. C. J., Benian, G. M., and Kemp, B. E. (1994) Insights into autoregulation from the crystal structure of twitchin kinase. *Nature* **369**, 581–584
32. Carmel, G., Leichus, B., Cheng, X., Patterson, S. D., Mirza, U., Chait, B. T., and Kuret, J. (1994) Expression, purification, crystallization, and preliminary X-ray analysis of casein kinase-1 from *Schizosaccharomyces pombe*. *J. Biol. Chem.* **269**, 7304–7309
33. Zheng, J., Knighton, D. R., Xuong, N. H., Taylor, S. S., Sowadski, J. M., and Ten Eyck, L. F. (1993) Crystal structures of the myristylated catalytic subunit of cAMP-dependent protein kinase reveal open and closed conformations. *Protein Sci.* **2**, 1559nd573
34. Knighton, D. R., Cadena, D. L., Zheng, J., Ten Eyck, L. F., Taylor, S. S., Sowadski, J. M., and Gill, G. N. (1993) Structural features that specify tyrosine kinase activity deduced from homology modeling of the epidermal growth factor receptor. *Proc. Natl. Acad. Sci. USA* **90**, 5001–5005
35. Taylor, S. S., Knighton, D. R., Zheng, J., Ten Eyck, L. F., and Sowadski, J. M. (1992) Structural framework of the protein kinase family. *Annu. Rev. Cell Biol.* **8**, 429–462
36. Taylor, S. S., Zheng, J., Radzio-Andzelm, E., Knighton, D. R., Ten Eyck, L. F., Sowadski, J. M., Herberg, F. W., and Yonemoto, W. (1993) cAMP-dependent protein kinase defines a family of enzymes. *Phil. Trans. R. Soc. London B* **340**, 315–324
37. Madhusudan, A., Trafny, E. A., Xuong, N. H., Adams, J. A., Ten Eyck, L. F., Taylor, S. S., and Sowadski, J. M. (1994) cAMP-dependent protein kinase: crystallographic insights into substrate recognition and phosphotransfer. *Protein Sci.* **3**, 176–187
38. Taylor, S. S., Radzio-Andzelm, E. (1994) Three protein kinase structures define a common motif. *Structure* **2**, 345–355
39. Steinberg, R. A., Cauthron, R. D., Symcox, M. M., and Shuntoh, H. (1993) Autoactivation of catalytic (Cα) subunit of cyclic AMP-dependent protein kinase by phosphorylation at threonine 197. *Mol. Cell. Biol.* **13**, 2332–2341
40. Veron, M., Radzio-Andzelm, E., Tsigelny, I., Ten Eyck, L. F., and Taylor, S. S. (1993) A conserved helix motif complements the protein kinase core. *Proc. Natl. Acad. Sci. USA* **90**, 10618–10622
41. Swofford, D. (1991) PAUP: Phylogenetic Analysis Using Parsimony, Version 3.1. Illinois Natural History Survey, Champaign, Illinois
42. Hardie, D. G., and Hanks, S. K. (1995) The Protein Kinase Factsbook, Academic Press, London
43. Pearson, R. B., Kemp, B. E. (1991) Protein kinase phosphorylation site sequences and consensus specificity motifs: tabulations. *Meth. Enzymol.* **200**, 62–81
44. Songyang, Z., Blechner, S., Piwnica-Worms, H., and Cantley, L. C. (1994) A novel oriented peptide library technique for determining optimal substrates of protein kinases. *Curr. Biol.* In press
45. Lindberg, R. A., Quinn, A. M., and Hunter, T. (1992) Dual-specificity protein kinases: will any hydroxyl do? *Trends Biochem. Sci.* **17**, 114–119
46. Koonin, E. V., Bork, P., and Sander, C. (1994) Yeast chromosome III: new gene functions *EMBO J.* **13**, 493–503
47. Johnston, M., Andrews, S., Brinkman, R., Cooper, J., Ding, H., Dover, J., Du, Z., Pavello, A., Fulton, L., Gattung, S., et al. (1994) Complete nucleotide sequence of *Saccharomyces cerevisiae* chromosome VIII. *Science* **265**, 2077–2082
48. Wilson, R., Ainscough, R., Anderson, K., Baynes, C., Berks, M., Bonfield, J ., Burton, J., Connell, M., Copsey, T., Cooper, J., et al. (1994) 2.2 Mb of contiguous nucleotide sequence from chromosome III of *C. elegans.*. *Nature* **368**, 32–38

Exhibit 8

58. M. L. Alexander, M. A. Johnson, W. C. Lineberger, *J. Chem. Phys.* **82**, 5288 (1985).
59. D. E. Hunton, M. Hofmann, T. G. Lindeman, C. R. Albertoni, A. W. Castleman, Jr., *ibid.*, p. 2884.
60. W. D. Reents, Jr., A. M. Mujsce, V. E. Bondybey, M. L. Mandich, *ibid.* **86**, 5568 (1987).
61. M. D. Morse, M. E. Geusic, J. R. Heath, R. E. Smalley, *ibid.* **83**, 2293 (1985).
62. E. K. Parks, K. Liu, S. C. Richtsmeier, L. G. Pobo, S. J. Riley, *ibid.* **82**, 5470 (1985).
63. S. A. Ruatta, L. Hanley, S. L. Anderson, *Chem. Phys. Lett.* **137**, 5 (1987).

64. R. D. Levine, R. B. Bernstein, Eds., *Molecular Reaction Dynamics and Chemical Reactivity* (Oxford Univ. Press, New York, 1987).
65. B. Koplitz *et al.*, *Faraday Discuss. Chem. Soc.* **82**, 125 (1986).
66. N. F. Scherer, L. R. Khundkar, R. B. Bernstein, A. H. Zewail, *J. Chem. Phys.* **87**, 1451 (1987).
67. S. Morgan and A. W. Castleman, Jr., *J. Am. Chem. Soc.* **109**, 2867 (1987).
68. D. K. Bohme, *NATO ASI Ser. C* **118**, 111 (1984).
69. Financial support by the Army Research Office, grant DAAG29-85-K-0215, the Department of Energy, grant DE-AC02-82ER60055, and the National Science Foundation, grant ATM-82-04010, is gratefully acknowledged.

The Protein Kinase Family: Conserved Features and Deduced Phylogeny of the Catalytic Domains

STEVEN K. HANKS, ANNE MARIE QUINN, TONY HUNTER

In recent years, members of the protein kinase family have been discovered at an accelerated pace. Most were first described, not through the traditional biochemical approach of protein purification and enzyme assay, but as putative protein kinase amino acid sequences deduced from the nucleotide sequences of molecularly cloned genes or complementary DNAs. Phylogenetic mapping of the conserved protein kinase catalytic domains can serve as a useful first step in the functional characterization of these newly identified family members.

T HE PROTEIN KINASES ARE A LARGE FAMILY OF ENZYMES, many of which mediate the response of eukaryotic cells to external stimuli (*1*, *2*). The number of unique members of the protein kinase family that have been described has recently risen exponentially (*3*) and now approaches 100. The surge in the number of known protein kinases has been largely due to the advent of gene cloning and sequencing techniques. Amino acid sequences deduced from nucleotide sequences are considered to represent protein kinases if they include certain key residues that are highly conserved in the protein kinase "catalytic domain."

Two different molecular approaches have been most instrumental in the isolation of novel protein kinase–encoding genes or cDNAs: (i) complementation or suppression of genetic defects in invertebrate regulatory mutants, and (ii) screening DNA libraries by using protein kinase genes as hybridization probes under low stringency conditions. Recently, an approach that uses degenerate oligonucleotides as probes has led to the identification of several novel putative

protein kinase genes and cDNAs (*4*, *5*). The oligonucleotide probes are designed to recognize target sequences that encode short amino acid stretches highly conserved in protein kinase catalytic domains.

In this article, we present an alignment of catalytic domain amino acid sequences from 65 different members of the protein kinase family, including many putative protein kinase sequences recently deduced from nucleotide sequence data. Based on this alignment, we first identify and discuss conserved features of the catalytic domains and then provide a visual display of the various intersequence relations through construction of a catalytic domain phylogenetic tree. Catalytic domains from protein kinases having similar modes of regulation or substrate specificities are found to cluster together within the tree. This clustering would appear to be of predictive value in the determination of the properties and function of novel protein kinases.

Catalytic Domain Amino Acid Sequences

Protein kinase catalytic domains range from 250 to 300 amino acid residues, corresponding to about 30 kD. Fairly precise boundaries for the catalytic domains have been defined through an analysis of conserved sequences (*1*, *6*, see below) as well as by assay of truncated enzymes (*7*, *8*). The location of the catalytic domain within the protein is not fixed but, in most single subunit enzymes it lies near the carboxyl terminus, the amino terminus being devoted to a regulatory role. In protein kinases having a multiple subunit structure, subunit polypeptides consisting almost entirely of catalytic domain are common. All protein kinases thus far characterized with regard to substrate specificity fall within one of two broad classes, serine/threonine–specific and tyrosine-specific. Although both classes of protein kinase have very similar catalytic domain primary structures, certain short amino acid stretches appear to characterize each class (*4*), and these regions can be used to predict whether a putative protein kinase will phosphorylate tyrosine or serine/threonine.

Members of the protein–serine/threonine kinase and protein-

S. K. Hanks is a senior research associate at the Molecular Biology Laboratory, Salk Institute for Biological Studies, Post Office Box 85800, San Diego, CA 92138. A. M. Quinn is a scientific applications programmer at the Biocomputing Center, Salk Institute for Biological Studies, Post Office Box 85800, San Diego, CA 92138. T. Hunter is a professor at the Molecular Biology and Virology Laboratory, Salk Institute for Biological Studies, Post Office Box 85800, San Diego, CA 92138.

tyrosine kinase families with reported catalytic domain amino acid sequences are listed in Tables 1 and 2, respectively. They are classified within the tables according to similarities in primary structure, based on deduced catalytic domain phylogeny. Included in the tables are all confirmed and putative protein kinases for which the catalytic domain sequence was available as of November 1987 (9). Presumed functional homologs from different vertebrate species are listed together. Presumed invertebrate functional homologs of protein kinases also found in vertebrates, however, are given

separate listings as a reflection of greater evolutionary distance and the possibility of functional divergence. The asterisks indicate protein kinases that have catalytic domains that are included in the amino acid sequence alignment. We will use the abbreviated names from the tables to refer to individual protein kinases.

Of the 45 unique vertebrate protein kinase family members included in Tables 1 and 2, 22 are serine/threonine-specific and 23 are tyrosine-specific. Fourteen of the vertebrate protein–serine/threonine kinases fall within one of the three subgroups that can be

Table 1. Protein–serine/threonine kinase family members.

A. Cyclic nucleotide–dependent subfamily
cAPK-α: cAMP-dependent protein kinase catalytic subunit, α form
    *-bovine cardiac muscle protein (26)
    -mouse S49 lymphoma cell cDNA (35)
cAPK-β: cAMP-dependent protein kinase catalytic subunit, β form
    *-bovine pituitary cDNA (36)
    -mouse S49 lymphoma cell cDNA (37)
SRA3: cAMP-dependent protein kinase from yeast, RAS suppressor
    *-Saccharomyces cerevisiae genomic DNA (38)
TPK1(PK25): cAMP-dependent protein kinase from yeast, type 1
    *-S. cerevisiae genomic DNA (39, 40)
TPK2: cAMP-dependent protein kinase from yeast, type 2
    *-S. cerevisiae genomic DNA (39)
TPK3: cAMP-dependent protein kinase from yeast, type 3
    *-S. cerevisiae genomic DNA (39)
cGPK: guanosine 3',5'-monophosphate (cGMP)–dependent protein kinase
    *-bovine lung protein (41)

B. Calcium-phospholipid–dependent subfamily
PKC-α: protein kinase C, α form
    *-bovine brain cDNA (42)
    -rabbit brain cDNA (43)
    -human brain cDNA (partial) (44)
PKC-β: protein kinase C, β form
    *-bovine brain cDNA (44)
    -rat brain cDNA (two splice forms) (45, 46)
    -rabbit brain cDNA (two splice forms) (43)
    -human brain cDNA (44)
PKC-γ: protein kinase C, γ form
    *-bovine brain cDNA (44)
    -rat brain cDNA (45)
    -human brain cDNA (44)
PKC-ε: protein kinase C, ε form
    -rat brain cDNA (RP16 clone) (partial) (46)
DPKC: Drosophila gene product related to protein kinase C
    *-D. melanogaster cDNA (47)

C. Calcium-calmodulin–dependent subfamily
CaMII-α: calcium-calmodulin–dependent protein kinase type II, α subunit
    *-rat brain cDNA (48)
CaMII-β: calcium-calmodulin–dependent protein kinase type II, β subunit
    *-rat brain cDNA (49)
PhK-γ: phosphorylase kinase, γ subunit
    *-rabbit skeletal muscle protein and cDNA (50)
    -mouse muscle cDNA (51)
MLCK-K: myosin light chain kinase, skeletal muscle
    *-rabbit skeletal muscle protein (52)
MLCK-M: myosin light chain kinase, smooth muscle
    *-chicken gizzard cDNA (53)
PSK-H1: putative protein-serine kinase
    *-human HeLa cell cDNA (4, 54)
PSK-C3: putative protein-serine kinase
    -human HeLa cell cDNA (partial) (4)

D. SNF1 subfamily
SNF1: "sucrose nonfermenting" mutant wild-type gene product
    *-S. cerevisiae genomic DNA (55)
nim1⁺: "new inducer of mitosis"; suppressor of cdc25 mutants
    *-Schizosaccharomyces pombe genomic DNA (56)
KIN1: putative yeast protein kinase
    *-Saccharomyces cerevisiae genomic DNA (5)
KIN2: putative yeast protein kinase related to KIN1
    *-S. cerevisiae genomic DNA (5)
E. CDC28-cdc2⁺ subfamily
CDC28: "cell-division-cycle" gene product in yeast
    *-S. cerevisiae genomic DNA (57)
cdc2⁺: "cell-division-cycle" gene product in yeast
    *-Schizosaccharomyces pombe genomic DNA (58)
CDC2Hs: human functional homolog of cdc2⁺
    *-human transformed cell line cDNA (33)
PSK-J3: putative protein kinase related to CDC28-cdc2⁺
    *-human HeLa cell cDNA (4, 59)
KIN28: putative protein kinase related to CDC28-cdc2⁺
    *-Saccharomyces cerevisiae genomic DNA (60)
F. Casein kinase subfamily
CKIIα: casein kinase II, α subunit
    -bovine lung protein (partial) (61)
DCKII: Drosophila casein kinase II, α subunit
    *-D. melanogaster cDNA (62)
G. Raf-Mos proto-oncogene subfamily
Raf: cellular homolog of oncogene products from 3611 murine sarcoma virus and Mill Hill 2 avian acute leukemia virus
    *-human fetal liver cDNA (63)
A-Raf: cellular oncogene product closely related to Raf
    *-human T cell cDNA (64)
    -mouse spleen cDNA (65)
PKS: cellular gene product closely related to Raf
    *-human fetal liver cDNA (66)
Mos: cellular homolog of oncogene product from Moloney murine sarcoma virus
    *-human placenta genomic DNA (67)
    -mouse NIH 3T3 cell genomic DNA (68)
    -rat 3Y1 cell genomic DNA (69)
H. STE7 subfamily
STE7: "sterile" mutant wild-type allele gene product
    *-S. cerevisiae genomic DNA (70)
PBS2: polymixin B antibiotic resistance gene product
    *-S. cerevisiae genomic DNA (71)
I. Family members with no close relatives
CDC7: "cell-division-cycle" gene product
    *-S. cerevisiae genomic DNA (72)
wee1⁺: "reduced size at division" mutant wild-type gene product
    *-Schizosaccharomyces pombe genomic DNA (73)
ran1⁺: "meiotic bypass" mutant wild-type allele gene product
    *-S. pombe genomic DNA (74)
PIM-1: putative transforming protein induced by murine leukemia virus integration
    *-mouse BALB/c cell genomic DNA (75)
HSVK: herpes simplex virus-US3 gene product
    *-herpes simplex virus genomic DNA (76)

*Protein kinases that have catalytic domains included in the amino acid sequence alignment.

classified according to their mode of regulation: cyclic nucleotide–dependent, calcium-phospholipid–dependent, and calcium-calmodulin–dependent. Two of the serine/threonine kinases, Mos and Raf (products of the c-*mos* and c-*raf* genes, respectively), are cellular homologs of transforming proteins encoded by the retroviral oncogenes. Other members of the serine/threonine group with demonstrated oncogenic potential are A-Raf (a distinct Raf-related member), and PIM-1 (a putative transforming protein activated by viral integration). Three vertebrate serine/threonine kinases (CDC2Hs, PSK-J3, and CKIIα) are closely related, by various degrees, to the yeast cell cycle control protein kinases CDC28 and cdc2⁺. A protein–serine/threonine kinase has been described in herpes simplex virus (HSVK) and, like the retroviral oncogenes, probably originated as a eukaryotic cellular sequence. The protein-tyrosine kinases can be further grouped as members of either the Src subfamily or one of three different growth factor receptor subfamilies. The protein-tyrosine kinases encoded by the c-*abl* and c-*fes/fps* genes may be considered distant members of the Src subfamily. At least nine of the protein-tyrosine kinase genes have been transduced

by retroviruses where they encode transforming proteins.

Twenty-five additional sequences listed in Tables 1 and 2 derive from invertebrate species. Eight are from *Drosophila*, one from nematode, and the other 16 are from the budding or fission yeasts. Many of the *Drosophila* protein kinases, as well as the nematode protein kinase, were identified by screening DNA libraries with probes from a vertebrate protein kinase gene or cDNA and thus are likely to represent functional homologs of the vertebrate enzymes. The *Drosophila* "sevenless" (7less) protein kinase and most of the yeast protein kinases were identified through molecular genetics. All of the yeast protein kinases identified to date fall within the serine/threonine–specific class, despite directed attempts to identify protein-tyrosine kinases in yeast (5). This observation, together with the fact that many of the protein-tyrosine kinase catalytic domains are components of growth factor receptor molecules, suggests that tyrosine specificity may have been a recent development in catalytic domain evolution, arising in conjunction with the acquisition of multicellularity and serving a role in cell-cell communication.

Table 2. Protein-tyrosine kinase family members.

A. **Src subfamily**
Src: cellular homolog of oncogene product from Rous avian sarcoma virus
  *-human fetal liver genomic DNA (77)
  -mouse brain cDNA; neuronal alternate splice form (78)
  -chicken genomic DNA (79)
  -*Xenopus laevis* ovary cDNA (partial) (80)
Yes: cellular homolog of oncogene product from Yamaguchi 73 avian sarcoma virus
  *-human embryo fibroblast cDNA (81)
Fgr: cellular homolog of oncogene product from Gardner-Rasheed feline sarcoma virus
  *-human genomic DNA (82)
  -human B lymphocyte cell line cDNA (amino terminus) (83)
FYN: putative protein-tyrosine kinase related to Fgr and Yes
  *-human fibroblast cDNA (84)
LYN: putative protein-tyrosine kinase related to LCK and Yes
  *-human placenta cDNA (85)
LCK: lymphoid cell protein-tyrosine kinase
  *-human (JURKAT) T cell leukemia line cDNA (86)
  -mouse (LSTRA) T cell lymphoma line cDNA (87)
HCK: hematopoietic cell putative protein-tyrosine kinase
  *-human placenta and peripheral leukocyte cDNAs (88)
Dsrc64: *Drosophila* gene product related to Src; polytene locus 64B
  *-*D. melanogaster* genomic DNA (89, 90)
Dsrc28: *Drosophila* gene product related to Src; polytene locus 28C
  *-*D. melanogaster* adult female cDNA (91)

B. **Abl subfamily**
Abl: cellular homolog of oncogene product from Abelson murine leukemia virus
  *-human fetal liver cDNA (92)
ARG: putative protein-tyrosine kinase related to Abl
  -human genomic DNA (partial) (93)
Dash: *Drosophila* gene product related to Abl
  *-*D. melanogaster* genomic DNA (90)
Nabl: nematode gene product related to Abl
  *-*Caenorhabditis elegans* genomic DNA (94)
Fes/Fps: cellular homolog of oncogene products from Gardner-Arnstein and Snyder-Theilen feline sarcoma viruses and Fujinami and PRCII avian sarcoma viruses
  *-human genomic DNA (95)
  -feline genomic DNA (96)
  -chicken genomic DNA (97)

C. **Epidermal growth factor receptor subfamily**
EGFR: epidermal growth factor receptor; cellular homolog of

oncogene product (v-Erb-B) from AEV-H avian erythroblastosis virus
  *-human placenta and A431 cell line cDNAs (98)
NEU: cellular oncogene product activated in induced rat neuroblastomas (also called ERB-B2 or HER2)
  *-human placenta and gastric cancer cell line cDNAs (99)
  -rat neuroblastoma cell line cDNA (100)
DER: *Drosophila* gene product related to EGFR
  *-*D. melanogaster* genomic DNA (101)

D. **Insulin receptor subfamily**
INS.R: insulin receptor
  *-human placenta cDNA (102)
IGF1R: insulin-like growth factor 1 receptor
  *-human placenta cDNA (103)
DILR: *Drosophila* gene product related to INS.R
  *-*D. melanogaster* embryo cDNA (104)
Ros: cellular homolog of oncogene product from UR2 avian sarcoma virus
  *-human placenta genomic DNA (105)
  -chicken genomic DNA (106), chicken kidney cDNA (107)
7less: *Drosophila* sevenless gene product essential for R7 photoreceptor cell development
  *-*D. melanogaster* eye imaginal disc cDNA (108)
TRK: colon carcinoma oncogene product activated by genetic recombination
  *-human tumor cell cDNA (109)
MET: N-methyl-N'-nitro-N-nitrosoguanidine (MNNG)–induced oncogene product
  *-human HOS cell line cDNA (110)

E. **Platelet-derived growth factor receptor subfamily**
PDGFR: platelet-derived growth factor receptor
  *-mouse NR6 fibroblast cell line cDNA (111)
CSF1R: colony-stimulating factor–type 1 receptor; cellular homolog of oncogene product (v-Fms) from McDonough feline sarcoma virus
  *-human placenta cDNA (112)
Kit: cellular homolog of oncogene product from Hardy-Zuckerman 4 feline sarcoma virus
  *-human placenta cDNA (113)
RET: cellular oncogene product activated by recombination
  *-human T cell lymphoma cDNA (114)

F. **Other receptor-like protein-tyrosine kinases**
TKR11: putative protein-tyrosine kinase
  -chicken genomic DNA (partial) (115)
TKR16: putative protein-tyrosine kinase
  -chicken genomic DNA (partial) (115)

*Protein kinases that have catalytic domains included in the amino acid sequence alignment.

Conserved Features of the Catalytic Domains

To compare primary structures, we have aligned catalytic domains from the 65 protein kinases marked by an asterisk in Tables 1 and 2 (Fig. 1). The 65 sequences represent each of the separate entries in the Tables except for six family members that are not included because their catalytic domain sequences have been only partially determined. The alignment was made by eye and is parsimonious in nature; the amount of gapping introduced into the sequences in order to optimize positional similarities was kept to a minimum. The alignment clearly demonstrates the overall similarity among the catalytic domains. The catalytic domains are not conserved uniformly but, rather, consist of alternating regions of high and low conservation. Eleven major conserved subdomains are evident (Fig. 1, I to XI), separated by regions of lower conservation wherein fall the larger gaps or inserts. Very large inserts (in excess of 60 residues) occur in CDC7 between subdomains VII and VIII and between subdomains X and XI, and in PDGFR, CSF1R, and Kit between subdomains V and VI. A similarity profile of the aligned catalytic domains provides a ready visualization of the subdomain structure (Fig. 2). Such an arrangement of alternating regions of high and low conservation is a common feature of homologous globular proteins (10) and gives some clues to higher order structure. The conserved subdomains must be important for catalytic function, either directly as components of the active site or indirectly by contributing to the formation of the active site through constraints imposed on secondary structure. The nonconserved regions, on the other hand, are likely to occur in loop structures, where folding allows the essential conserved regions to come together.

Highly conserved individual amino acids within the catalytic domains are expected to play important roles in catalysis. We will refer to amino acid positions using the residue numbering for bovine adenosine 3',5'-monophosphate (cAMP)–dependent protein kinase catalytic subunit, α form (cAPK-α, Fig. 1). Nine positions in the alignment contain the identical amino acid residue in each of the 65 sequences. These invariant residues correspond to cAPK-α: Gly[52], Lys[72], Glu[91], Asp[166], Asn[171], Asp[184], Gly[186], Glu[208], and Arg[280]. An additional five positions contain the identical amino acid in all but one of the sequences: Gly[50], Val[57], Phe[185], Asp[220], and Gly[225]. Many of these most highly conserved residues directly participate in adenosine triphosphate (ATP) binding and phosphotransfer.

The consensus Gly-X-Gly-X-X-Gly, found in many nucleotide binding proteins in addition to the protein kinases (11), is found in subdomain I, very near the catalytic domain amino terminus. The invariant or nearly invariant residues corresponding to cAPK-α Gly[50] and Gly[52] fall within this consensus. Only two positions on the amino-terminal side of this consensus show conservation throughout the protein kinase family; hydrophobic residues occupy positions one and seven upstream from the first glycine in the consensus. The amino terminus of some catalytic domain polypeptides lies as close as ten residues from the first conserved glycine. A model for the ATP-binding site of v-Src (12), based on the three-dimensional structures from other nucleotide binding proteins, shows the Gly-X-Gly-X-X-Gly residues forming an elbow around the nucleotide, with the first glycine in contact with the ribose moiety and the second glycine lying near the terminal pyrophosphate. A nearly invariant valine residue lies within subdomain I, located just two positions on the carboxyl-terminal side of the Gly-X-Gly-X-X-Gly consensus (Val[57] for cAPK-α) and may contribute to the positioning of the conserved glycines.

In subdomain II lies an invariant lysine, corresponding to cAPK-α Lys[72], that is certainly the best characterized catalytic domain residue. This lysine appears to be directly involved in the phospho-transfer reaction, possibly mediating proton transfer (13). In cAPK-α (14), v-Src (15), and EGFR (16), Lys[72] or its equivalent reacts with the ATP analog p-fluorosulfonyl 5'-benzoyl adenosine, thereby inhibiting enzyme activity. Site-directed mutagenesis techniques have been used to substitute alternate amino acids at this position in v-Src (13, 17), v-Mos (18), v-Fps (19), EGFR (20), INS.R (21), and PDGFR (22). All substitutions, including arginine, result in loss of protein kinase activity. In all but three of the aligned sequences, an alanine is present two positions on the amino-terminal side of the invariant lysine in subdomain II. The invariant lysine lies 14 to 23 residues downstream of the last conserved glycine in subdomain I, but no mutations have been made to test whether this spacing is critical.

The central core of the catalytic domain, the region with greatest frequency of highly conserved residues, consists of subdomains VI through IX. The invariant or nearly invariant residues in subdomain VI (corresponding to Asp[166] and Asn[171]) and subdomain VII (corresponding to Asp[184], Phe[185], and Gly[186]) also have been implicated in ATP binding. These residues are part of a feature found in a number of bacterial phosphotransferases that use ATP as phosphate donor (23). The aspartic acid residues corresponding to cAPK-α Asp[166] and Asp[184] may interact with the phosphate groups of ATP through Mg^{2+} salt bridges (23). The triplet corresponding to Asp[184]-Phe[185]-Gly[186] in subdomain VII is of further interest in that it represents the most highly conserved short stretch in the catalytic domains. It is flanked for two positions on either side by hydrophobic or near-neutral residues.

Subdomain VIII contains the consensus triplet Ala-Pro-Glu, a conserved feature often mentioned as a key protein kinase catalytic domain indicator (1). The invariant residue corresponding to cAPK-α Glu[208] contributes to the Ala-Pro-Glu consensus. In addition to the conservation of these residues, several other lines of evidence implicate this region as important in catalysis. Mutagenesis studies have shown that each residue in the Ala-Pro-Glu consensus is required for activity of v-Src (24). Other studies have provided evidence that this consensus lies very near the catalytic site. An affinity peptide substrate analog reacts with cAPK-α Cys[199], thereby inhibiting enzyme activity (25). Also, sites of autophosphorylation found in many protein-tyrosine kinases (1) as well as cAMP-dependent protein kinase [Thr[197] (26)] lie within 20 residues upstream of the Ala-Pro-Glu consensus. The role of this autophosphorylation site is not entirely settled, but for several protein-tyrosine kinases there is evidence that phosphorylation of this site leads to increased catalytic activity (27). Autophosphorylation may result in a conformational change that allows better access of exogenous substrates to the active site.

Subdomains VI and VIII are of additional interest in that they contain residues that are specifically conserved in either the protein-serine/threonine or the protein-tyrosine kinases and, as such, may play a role in recognition of the correct hydroxyamino acid. The most striking indicator of amino acid specificity is found in subdomain VI, lying between the invariant residues corresponding to cAPK-α Asp[166] and Asn[171]; two of the residues implicated in ATP binding. The consensus Asp-Leu-Lys-Pro-Glu-Asn in this region is a strong indicator of serine/threonine specificity, whereas the protein-tyrosine kinase consensus is either Asp-Leu-Arg-Ala-Ala-Asn (for the vertebrate members of the Src subfamily) or Asp-Leu-Ala-Ala-Arg-Asn (for all others). Another such region is found in subdomain VIII and lies immediately on the amino-terminal side of the Ala-Pro-Glu consensus. This region is highly conserved among the protein-tyrosine kinases with a more limited conservation among the protein-serine/threonine kinases. The protein-tyrosine kinase consensus through this region is Pro-Ile/Val-Lys/Arg-Trp-Thr/Met-Ala-Pro-Glu while the protein-serine/threonine kinase

consensus is Gly-Thr/Ser-X-X-Tyr/Phe-X-Ala-Pro-Glu. These regions in subdomains VI and VIII that indicate substrate specificity have been targeted for the design of degenerate oligonucleotide probes for use in screening cDNA libraries to identify novel members of both the protein–serine/threonine (4) and protein-tyrosine (28) kinase families.

To date, no evidence has been reported concerning the possible functions of residues in conserved subdomains III, IV, V, IX, X, and XI. Subdomain IX contains a very well conserved short stretch that includes the nearly invariant residues corresponding to Asp^{220} and Gly^{225}. Subdomains III and XI each contain an invariant residue, corresponding to Glu^{91} and Arg^{280}. The latter or its equivalent must lie very near the catalytic domain carboxyl terminus. Arginine residues occupying this position reside just 16 residues upstream from both the CDC28 and HSVK polypeptide carboxyl termini, and just 19 residues upstream from both the Mos and Fes carboxyl termini. Deletion analysis of v-Src places the carboxyl terminus of the catalytic domain of the protein-tyrosine kinases at a conserved hydrophobic residue ten residues downstream of this arginine (8). The point mutation conferring temperature sensitivity in some cdc28 mutants replaces this conserved arginine with glutamine (29).

A leap in our understanding of the functional roles of the conserved catalytic domain residues will come with the solution of a crystal structure for one of the protein kinase catalytic domains. The similarities in primary strucure should carry over to the higher order structure and catalytic mechanism as well. Other investigators have been making progress toward the solution of the three-dimensional structure of cAPK-α (30).

Catalytic Domain Phylogeny

Amino acid sequence alignments can be used to deduce phylogenetic relationships (31). We have used the alignment data from Fig. 1 to construct a phylogenetic tree of the protein kinase catalytic domains (Fig. 3). All 65 of the sequences in the alignment are included in the tree. They derive from both vertebrate and invertebrate sources and, in some cases, presumed functional homologs from both vertebrate and invertebrate sources are represented. The tree, therefore, reflects catalytic domain evolution stemming from gene duplication events (for example, when the vertebrate, mostly human, sequences are compared), speciation events (when vertebrate and invertebrate functional homologs are compared), or both.

The tree reveals a relation between catalytic domain sequence and certain biochemical properties; catalytic domains from protein kinases having similar modes of regulation or substrate specificities tend also to have similar primary structures and cluster together within the tree. Five major branch clusters are present in the tree: (i) protein-tyrosine kinases, (ii) cyclic nucleotide– and calcium-phospholipid–dependent protein kinases, (iii) calcium-calmodulin–dependent protein kinases, (iv) protein kinases closely related to SNF1, and (v) protein kinases closely related to CDC28. These major clusters account for all but 12 of the 65 sequences included in the tree. Generally, a sequence found within one of these clusters shares in excess of 35% identical amino acids with each of the other sequences in the cluster, whereas the catalytic domain sequences that do not map within the same cluster have identities in the range of 20 to 25%.

The most highly populated cluster contains all 27 confirmed or putative protein-tyrosine kinases. The large number of protein-tyrosine kinases probably reflects the intense research effort devoted to this group, rather than a true indication of their abundance relative to the protein–serine/threonine kinases. Branches leading to the Src subfamily and to each of the three receptor subfamilies

diverge from the main line at about the same point. In light of the oncogenic potential of many of the protein-tyrosine kinases, it is of interest that the protein–serine/threonine kinases having the least divergence from this group include Raf and Mos, cellular homologs of retroviral oncogene products. However, another potentially oncogenic protein–serine/threonine kinase, PIM-1, is not closely related to the protein-tyrosine kinases.

The next most populous cluster in the tree includes two separate subfamilies that can be classified according to their mode of regulation: the cyclic nucleotide–dependent protein kinases and the calcium-phospholipid–dependent protein kinases. The similarities in the mode of regulation of the members of these two subfamilies, namely, activation by "second messengers" released in response to ligand binding at the cell surface, may be a reflection of their recent evolutionary divergence.

The third major catalytic domain cluster contains the subfamily of protein kinases that have activities regulated by calmodulin. The calmodulin-dependent cluster falls near the cyclic nucleotide– and calcium-phospholipid–dependent cluster. All members of the calmodulin-dependent subfamily have a calmodulin binding domain, characterized by a high proportion of basic amino acid residues and having a propensity for formation of an amphiphilic α helix, residing outside the catalytic domain. (Note that the calmodulin binding domain sequences were not included in the phylogenetic analysis.) The different protein kinases thus far described as being regulated by calmodulin, therefore, appear to have diverged from a common ancestor after acquisition of the calmodulin binding domain. The mapping of the putative protein kinase PSK-H1 within this cluster predicts that this enzyme will also prove to be regulated by calmodulin.

Also mapping near the cyclic nucleotide– and calcium-phospholipid–dependent protein kinases is a small cluster composed of four protein kinases recently identified in the budding or fission yeasts; SNF1, niml$^+$, KIN1, and KIN2. Whether these protein kinases

Fig. 1. Multiple amino acid sequence alignment of 65 protein kinase catalytic domains. The first 38 sequences derive from protein–serine/threonine kinases (indicated by asterisks in Table 1) and the remaining 27 sequences in the alignment are from protein-tyrosine kinases (indicated by asterisks in Table 2). cAPK-α and Src have been chosen as prototype protein–serine/threonine and protein-tyrosine kinases, respectively; their catalytic domain sequences are numbered to indicate residue position from the polypeptide amino terminus. (Although the human Src sequence is shown, the numbering is actually taken from the chicken Src sequence to maintain established convention). The number of additional amino- and carboxyl-terminal flanking residues lying outside the catalytic domains are shown at the beginning and end, respectively, of each sequence. In several cases the sequences have not been determined through to the polypeptide amino or carboxyl termini; for these, the number of determined residues is given followed by a plus (+) sign. An asterisk (*) at the beginning or end of a sequence indicates that no additional flanking residues are contained in the polypeptide. Gaps, represented by dashes, were introduced into the sequences to optimize the alignment. In six cases, long insert segments have been excluded from the alignment to shorten the figure. The positions and lengths of the excluded inserts within the alignment are indicated by numbers within braces (for example, {-48-}); the excluded gap positions in the other sequences that correspond to these long inserts are shown as double slashes (//). Residues conserved in 62 or more of the 65 sequences are shown as white letters in black boxes. Positions where residues of similar structure are conserved in 63 or more sequences are shown in shaded boxes. Structurally similar groupings used for this purpose are nonpolar chain R groups (M, L, I, V, and C); aromatic or ring-containing R groups; (F, Y, W, and H); small R groups with near neutral polarity (A, G, S, T, and P); acidic and uncharged polar R groups (D, E, N, and Q); and basic polar R groups (K, R, and H). The single-letter amino acid code is used (A, alanine; C, cysteine; D, aspartic acid; E, glutamic acid; F, phenylalanine; G, glycine; H, histidine; I, isoleucine; K, lysine; L, leucine; M, methionine; N, asparagine; P, proline; Q, glutamine; R, arginine; S, serine; T, threonine; V, valine; W, tryptophan; and Y, tyrosine). Roman numerals at bottom indicate conserved subdomains.

IX

X

XI

CAPK-α
CAPK-β
SRA3
TPK1
TPK2
TPK3
CGPK

PKC-α
PKC-β
PKC-γ
DPKC

CaMII-α
CaMII-β
PHK-γ
MLCK-K
MLCK-N
PSK-H1

SMK1
KIN1
KIN2

CDC28
cdc2+
CDC2Hs
PSK-J3
KIN28

DDK1
CDC7
STE7
PBS2
wee1+
Mos

Raf
A-Raf
Pks
Mos

Src
Yes
Fgr
Lyn
Hck
Lck
D-src4
D-src28

Abl
Dash
Nabl
Fes

EGFR
NEU
DER

INS.R
IGF1R
DILR
Ros
7less
.RR
NEI

PDGFR
CSF1R
Kit
RET

Fig. 2. Similarity profile of protein kinase catalytic domains. For each position in the alignment shown in Fig. 1, a relative similarity score was determined based on the "structure-genetic" scoring matrix (116) for amino acid similarities. Similarity scores were calculated as the sum of all possible pairwise comparisons between the individual amino acids at each position and expressed as the percentage of the highest possible score (that is, the score obtained when an identical residue occupies the position in all 65 aligned sequences). To smooth out the curve, a 9-position running average of the relative scores was determined, and every third position was plotted. Positions that contain gaps for ten or more of the sequences were not included in the profile; however, the locations of the major gap sites are indicated by breaks in the curve. The mean relative score for all the positions included in the profile is 66 with a standard deviation of 14.9. Relative similarity scores obtained when the catalytic domain sequences were randomly scrambled had a mean of 47 and standard deviation of 1.85. Roman numerals indicate conserved subdomains.
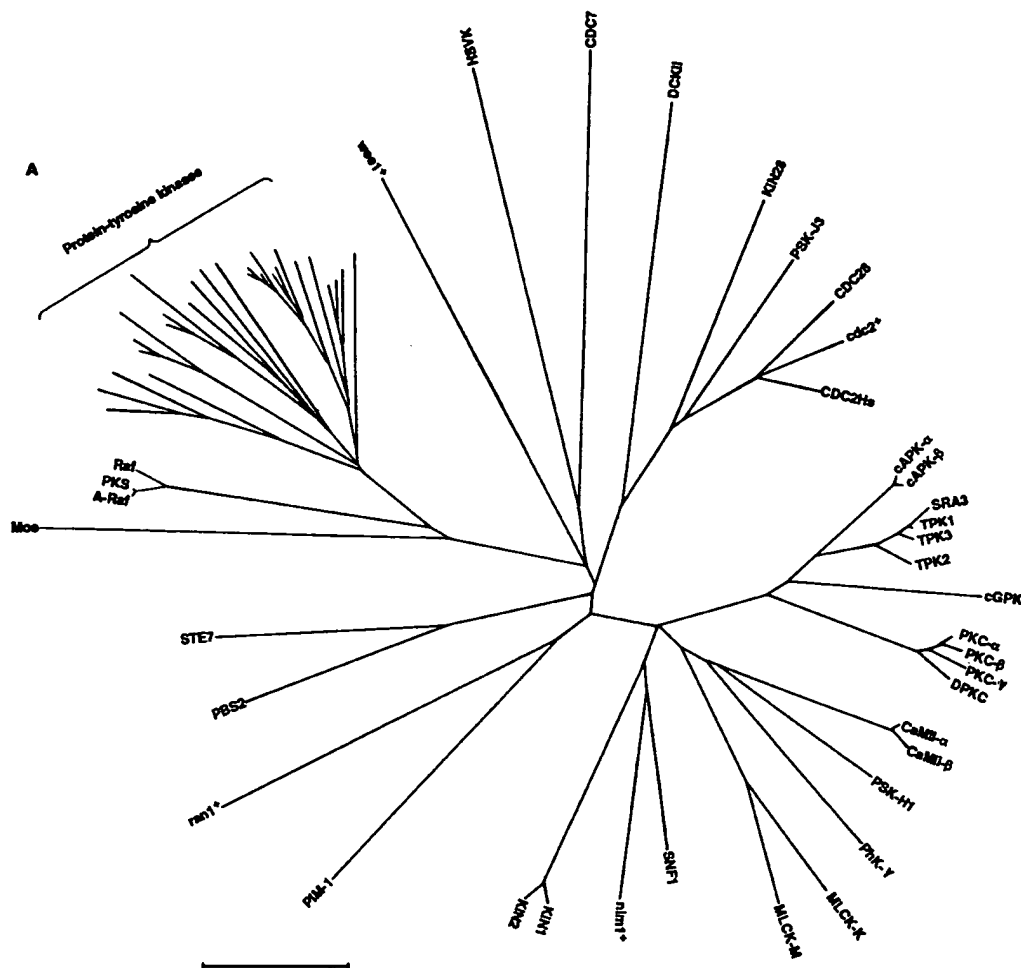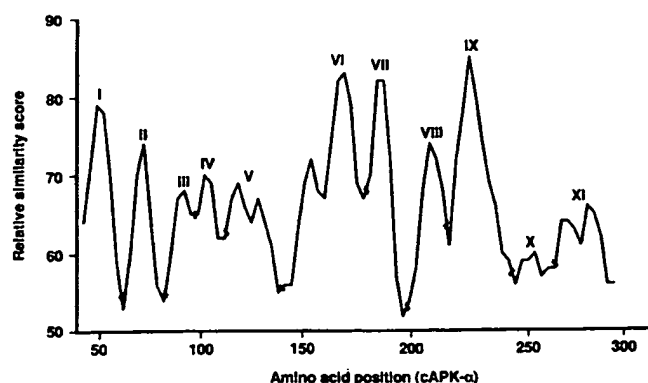

Fig. 3. Deduced phylogeny of protein kinase catalytic domains. The phylogenetic tree was constructed from the multiple alignment shown in Fig. 1. The tree-building concept of Fitch and Margoliash (117) was used as implemented by Feng and Doolittle (118). Briefly, similarity scores were obtained for all possible pairwise comparisons and transformed into a difference matrix from which branch order and length were determined. Programs were run on a VAX-785 computer equipped with 40 megabytes physical memory under virtual memory operating system (VMS). Systems limitations required that the branch lengths for the protein–serine/threonine and protein-tyrosine kinases be calculated separately, and the tree shown is thus a composite of these two determinations. The position of the protein-tyrosine kinase cluster was determined by including two protein-tyrosine kinases (Src and EGFR) in the protein–serine/threonine kinase tree construction. The individual sequences are indicated by the abbreviated names in Tables 1 and 2. The protein-tyrosine kinases are not labeled in (**A**), but are shown in the cluster enlargement in (**B**). The tree is shown "unrooted" in (A) as the branches are all measured relative to one another with no outside reference point. The scale bars represent a branch length corresponding to a relative difference score of 25. The tree depicted is likely to underestimate distances between the least related members of the family, particularly since the alignment used in its construction is parsimonious.
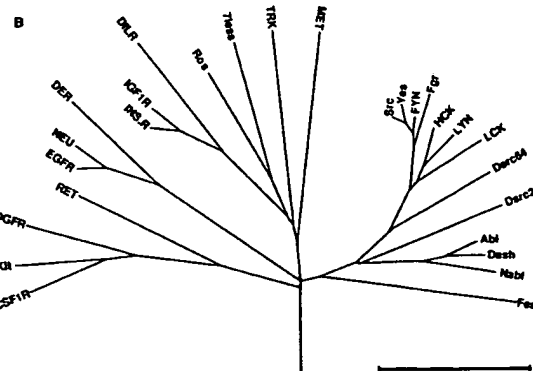
have similar modes of regulation remains to be determined. KIN1 and KIN2 were identified through screening a *Saccharomyces cerevisiae* DNA library with probes designed to recognize sequences characteristic of protein-tyrosine kinases and, as such, have been suggested to represent "structural mosaics" with some features of catalytic domain structure more indicative of the protein-tyrosine kinases than the protein–serine/threonine kinases (5). The deduced phylogeny of KIN1 and KIN2, however, does not suggest a close evolutionary relationship with protein-tyrosine kinases. In fact, the probe target used to identify KIN1 and KIN2 encodes the stretch of amino acids corresponding to cAPK-α Asp^{220}-Gly^{225} in conserved subdomain IX, a region of high conservation in all of the catalytic domains regardless of substrate specificity.

The subfamily related to CDC28 includes functional homologs from three widely divergent species: CDC28 from the budding yeast *S. cerevisiae*, cdc2$^+$ from the fission yeast *Schizosaccharomyces pombe*, and human CDC2Hs. Functional homology was demonstrated by heterologous complementation of conditional mutants defective in cell cycle progression (32, 33). The other two sequences mapping within this cluster are putative protein kinases identified in *Saccharomyces cerevisiae* (KIN28) and human HeLa cells (PSK-J3). The members of this cluster are also distinguished by the small sizes of the catalytic domain-containing polypeptides, suggesting their activities may be regulated through association with other polypeptides in a holoenzyme complex. Indeed, support for this notion has been obtained for cdc2$^+$ (34).

Perspectives

The tremendous diversity of the protein kinase family is just now beginning to be appreciated. Most of the catalytic domain sequences referenced in Tables 1 and 2 were reported within the past 2 years. With continued characterizations of regulatory mutants in invertebrates, along with the recent development of new hybridization approaches for the identification of DNA clones that encode novel protein kinase catalytic domains, it is likely that the rate of discovery will continue to accelerate through the next several years. The difficult tasks will be to confirm protein kinase activities for the newly identified family members and to elucidate their functional roles. Clues to function may come through an analysis of catalytic domain primary structure and subsequent phylogenetic mapping. A catalytic domain that has only limited divergence from another, better characterized, member of the family can be expected to play a similar role in cellular physiology. Further clues are likely to come from an inspection of amino acid sequences lying outside the catalytic domain where residues involved in enzyme regulation may be found.

REFERENCES AND NOTES

1. T. Hunter and J. A. Cooper, in *The Enzymes*, P. D. Boyer and E. G. Krebs, Eds. (Academic Press, Orlando, FL, 1986), vol. 17, pp. 191–246.
2. A. M. Edelman, D. K. Blumenthal, E. G. Krebs, *Annu. Rev. Biochem.* 56, 567 (1987).
3. T. Hunter, *Cell* 50, 823 (1987).
4. S. K. Hanks, *Proc. Natl. Acad. Sci. U.S.A.* 84, 388 (1987).
5. D. E. Levin, C. I. Hammond, R. O. Ralston, J. M. Bishop, *ibid.*, p. 6035.
6. W. C. Barker and M. O. Dayhoff, *ibid.* 79, 2836 (1982).
7. A. D. Levinson, S. A. Courtneidge, J. M. Bishop, *ibid.* 78, 1624 (1981); J. S. Brugge and D. Darrow, *J. Biol. Chem.* 259, 4550 (1984); J. Y. J. Wang and D. Baltimore, *ibid.* 260, 64 (1985); R. J. Bold and D. J. Donoghue, *Mol. Cell. Biol.* 5, 3131 (1985); I. Sadowski and T. Pawson, *Oncogene* 1, 181 (1987).
8. V. W. Wilkerson, D. L. Bryant, J. T. Parsons, *J. Virol.* 55, 314 (1985); P. Yaciuk and D. Shalloway, *Mol. Cell. Biol.* 6, 2807 (1986).
9. We have attempted to include all catalytic domain sequence reports except those encoded by retroviruses. Reports of partial sequences are included only if a complete catalytic domain sequence is not available. We have chosen not to include retroviral oncogene product sequences since the catalytic domains from

this group are effectively represented by their closely related cellular counterparts. A listing of references for retroviral protein kinase sequences can be found in *Molecular Biology of Tumor Viruses: RNA Tumor Viruses* [R. A. Weiss, N. Teich, H. Varmus, J. Coffin, Eds. (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, ed. 2, 1985)]. Since the time these tables were compiled, complete sequences for six additional members of the protein kinase family have been published: (i) EPH, a novel receptor-like protein-tyrosine kinase [H. Hirai, Y. Maru, K. Hagiwara, J. Nishida, F. Takaku, *Science* 238, 1717 (1987)]; (ii) TKL, a novel member of the Src subfamily [K. Strebhardt, J. I. Mullins, C. Bruck, H. Rübsamen-Waigmann, *Proc. Natl. Acad. Sci. U.S.A.* 84, 8778 (1987)]; (iii) *ninaC* protein, a *Drosophila* gene product essential for normal photoreceptor cell function [C. Montell and G. M. Rubin, *Cell* 52, 757 (1988)]; (iv) *nimA*, a cell cycle control protein kinase from *Aspergillus* [S. A. Osmani, R. T. Pu, N. R. Morris, *Cell* 53, 237 (1988)]. Further, in *Drosophila*, a full-length sequence of cAMP- and a partial sequence of cGMP-dependent protein kinase catalytic domain have been reported [J. L. Foster, G. C. Higgins, F. R. Jackson, *J. Biol. Chem.* 263, 1676 (1988)] as have catalytic domain sequences from a number of vertebrate protein kinases previously reported from other vertebrate species (not cited); (v) byr1$^+$, a suppressor of sporulation defects in *Schizosaccharomyces pombe* [S. A. Nadin-Davis and A. Nasim, *EMBO J.* 7, 985 (1988)]; and (vi) GCN2, a protein kinase essential for translational derepression of GCN4 mRNA in *Saccharomyces cerevisiae* [I. Roussou, G. Thireos, B. M. Hauge, *Mol. Cell. Biol.* 8, 2132 (1988)].
10. C. Chothia and A. M. Lesk, *EMBO J.* 5, 823 (1986).
11. R. K. Wierenga and W. G. J. Hol, *Nature* 302, 842 (1983).
12. M. J. E. Sternberg and W. R. Taylor, *FEBS Lett.* 175, 387 (1984).
13. M. P. Kamps and B. M. Sefton, *Mol. Cell. Biol.* 6, 751 (1986).
14. M. J. Zoller, N. C. Nelson, S. S. Taylor, *J. Biol. Chem.* 256, 10837 (1981).
15. M. P. Kamps, S. S. Taylor, B. M. Sefton, *Nature* 310, 589 (1984).
16. M. W. Russo, T. J. Lukas, S. Cohen, J. V. Staros, *J. Biol. Chem.* 260, 5205 (1985).
17. M. A. Snyder, J. M. Bishop, J. P. McGrath, A. D. Levinson, *Mol. Cell. Biol.* 5, 1772 (1985).
18. M. Hannink and D. J. Donoghue, *Proc. Natl. Acad. Sci. U.S.A.* 82, 7894 (1985).
19. G. Weinmaster, M. J. Zoller, T. Pawson, *EMBO J.* 5, 69 (1986).
20. W. S. Chen *et al.*, *Nature* 328, 820 (1987); A. M. Honegger *et al.*, *Cell* 51, 199 (1987).
21. C. K. Chou *et al.*, *J. Biol. Chem.* 262, 1842 (1987).
22. L. T. Williams, personal communication.
23. S. Brenner, *Nature* 329, 21 (1987).
24. D. Bryant and J. T. Parsons, *J. Virol.* 45, 1211 (1983); *Mol. Cell. Biol.* 4, 862 (1984).
25. H. N. Bramson *et al.*, *J. Biol. Chem.* 257, 10575 (1982).
26. S. Shoji *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* 78, 848 (1981); S. Shoji, L. H. Ericsson, K. A. Walsh, E. H. Fischer, K. Titani, *Biochemistry* 22, 3702 (1983).
27. G. Weinmaster, M. J. Zoller, M. Smith, E. Hinze, T. Pawson, *Cell* 37, 559 (1984); G. Weinmaster and T. Pawson, *J. Biol. Chem.* 261, 328 (1986); T. E. Kmiecik and D. Shalloway, *Cell* 49, 65 (1987); H. Piwnica-Worms, K. B. Saunders, T. M. Roberts, A. E. Smith, S. H. Cheng, *ibid.*, p. 75; R. Herrera and O. M. Rosen, *J. Biol. Chem.* 261, 11980 (1986); L. Ellis *et al.*, *Cell* 45, 721 (1986).
28. R. A. Lindberg, unpublished data.
29. A. Lörincz and S. I. Reed, *Mol. Cell. Biol.* 6, 4099 (1986).
30. J. M. Sowadski, N. h. Xuong, D. Anderson, S. S. Taylor, *J. Mol. Biol.* 182, 617 (1984).
31. R. F. Doolittle, in *The Proteins*, H. Neurath and R. L. Hill, Eds. (Academic Press, New York, ed. 3, 1979), vol. 4, pp. 1–118.
32. D. Beach, B. Durkacz, P. Nurse, *Nature* 300, 706 (1982); R. Booher and D. Beach, *Mol. Cell. Biol.* 6, 3523 (1986).
33. M. G. Lee and P. Nurse, *Nature* 327, 31 (1987).
34. L. Brizuela, G. Draetta, D. Beach, *EMBO J.* 6, 3507 (1987).
35. M. D. Uhler *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* 83, 1300 (1986).
36. M. O. Showers and R. A. Maurer, *J. Biol. Chem.* 261, 16288 (1986).
37. M. D. Uhler, J. C. Chrivia, G. S. McKnight, *ibid.*, p. 15360.
38. J. F. Cannon and K. Tatchell, *Mol. Cell. Biol.* 7, 2653 (1987).
39. T. Toda, S. Cameron, P. Sass, M. Zoller, M. Wigler, *Cell* 50, 277 (1987).
40. J. Liszewicz, A. Godany, H.-H. Forster, H. Kuntzel, *J. Biol. Chem.* 262, 2549 (1987).
41. K. Takio *et al.*, *Biochemistry* 23, 4207 (1984).
42. P. J. Parker *et al.*, *Science* 233, 853 (1986).
43. S. Ohno *et al.*, *Nature* 325, 161 (1987).
44. L. Coussens *et al.*, *Science* 233, 859 (1986).
45. J. L. Knopf *et al.*, *Cell* 46, 491 (1986).
46. G. M. Housey, C. A. O'Brian, M. D. Johnson, P. Kirschmeier, I. B. Weinstein, *Proc. Natl. Acad. Sci. U.S.A.* 84, 1065 (1987).
47. A. Rosenthal *et al.*, *EMBO J.* 6, 433 (1987).
48. R. M. Hanley *et al.*, *Science* 237, 293 (1987); C. R. Lin *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* 84, 5962 (1987).
49. M. K. Bennett and M. B. Kennedy, *Proc. Natl. Acad. Sci. U.S.A.* 84, 1794 (1987).
50. E. M. Reimann *et al.*, *Biochemistry* 23, 4185 (1984); E. F. da Cruz e Silva and P. T. W. Cohen, *FEBS Lett.* 220, 36 (1987).
51. J. S. Chamberlain, P. VanTuinen, A. A. Reeves, B. A. Philip, C. T. Caskey, *Proc. Natl. Acad. Sci. U.S.A.* 84, 2886 (1987).
52. K. Takio *et al.*, *Biochemistry* 24, 6028 (1985).
53. V. Guerriero, Jr., M. A. Russo, N. J. Olson, J. A. Putkey, A. R. Means, *ibid.* 25, 8372 (1986).
54. The entire catalytic domain sequence has been determined by J. R. Woodgett (unpublished data).
55. J. L. Celenza and M. Carlson, *Science* 233, 1175 (1986).

56. P. Russell and P. Nurse, *Cell* **49**, 569 (1987).
57. A. T. Lörincz and S. I. Reed, *Nature* **307**, 183 (1984).
58. J. Hindley and G. A. Phear, *Gene* **31**, 129 (1984).
59. The entire catalytic domain sequence has been determined by S. K. Hanks (unpublished data).
60. M. Simon, B. Seraphin, G. Faye, *EMBO J.* **5**, 2697 (1986).
61. K. Takio *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **84**, 4851 (1987).
62. A. Saxena, R. Padmanabha, C. V. C. Glover, *Mol. Cell. Biol.* **7**, 3409 (1987).
63. T. I. Bonner *et al.*, *Nucleic Acids Res.* **14**, 1009 (1986).
64. T. W. Beck, M. Huleihel, M. Gunnell, T. I. Bonner, U. R. Rapp, *ibid.* **15**, 595 (1987).
65. M. Huleihel *et al.*, *Mol. Cell. Biol.* **6**, 2655 (1986).
66. G. E. Mark, T. W. Seeley, T. B. Shows, J. D. Mountz, *Proc. Natl. Acad. Sci. U.S.A.* **83**, 6312 (1986).
67. R. Watson, M. Oskarsson, G. F. Vande Woude, *ibid.* **79**, 4078 (1982).
68. C. Van Beveren *et al.*, *Nature* **289**, 258 (1981).
69. F. A. Van der Hoorn and J. Firzlaff, *Nucleic Acids Res.* **12**, 2147 (1984).
70. M. A. Teague *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **83**, 7371 (1986).
71. G. Boguslawski and J. O. Polazzi, *ibid.* **84**, 5848 (1987).
72. M. Patterson, R. A. Sclafani, W. L. Fangman, J. Rosamond, *Mol. Cell. Biol.* **6**, 1590 (1986).
73. P. Russell and P. Nurse, *Cell* **49**, 559 (1987).
74. M. McLeod and D. Beach, *EMBO J.* **5**, 3665 (1986).
75. G. Selten *et al.*, *Cell* **46**, 603 (1986).
76. D. J. McGeoch and A. J. Davison, *Nucleic Acids Res.* **14**, 1765 (1986).
77. S. K. Anderson, C. P. Gibbs, A. Tanaka, H.-J. Kung, D. J. Fujita, *Mol. Cell. Biol.* **5**, 1122 (1985); A. Tanaka *et al.*, *ibid.* **7**, 1978 (1987).
78. R. Martinez, B. Mathey-Prevot, A. Bernards, D. Baltimore, *Science* **237**, 411 (1987).
79. T. Takeya and H. Hanafusa, *Cell* **32**, 881 (1983).
80. R. E. Steele, *Nucleic Acids Res.* **13**, 1747 (1985).
81. J. Sukegawa *et al.*, *Mol. Cell. Biol.* **7**, 41 (1987).
82. M. Nishizawa *et al.*, *ibid.* **6**, 511 (1986); R. C. Parker, G. Mardon, R. V. Lebo, H. E. Varmus, J. M. Bishop, *ibid.* **5**, 831 (1985).
83. K. Inoue *et al.*, *Oncogene* **1**, 301 (1987).
84. K. Semba *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **83**, 5459 (1986); T. Kawakami, C. Y. Pennington, K. C. Robbins, *Mol. Cell. Biol.* **6**, 4195 (1986).
85. Y. Yamanashi *et al.*, *Mol. Cell. Biol.* **7**, 237 (1987).
86. J. M. Trevillyan *et al.*, *Biochim. Biophys. Acta* **888**, 286 (1986); Y. Koga *et al.*, *Eur. J. Immunol.* **16**, 1643 (1986).
87. J. D. Marth, R. Peet, E. G. Krebs, R. M. Perlmutter, *Cell* **43**, 393 (1985); A. F. Voronova and B. M. Sefton, *Nature* **319**, 682 (1986).
88. N. Quintrell *et al.*, *Mol. Cell. Biol.* **7**, 2267 (1986); S. F. Ziegler, J. D. Marth, D. B. Lewis, R. M. Perlmutter, *ibid.*, p. 2276.
89. M. A. Simon, B. Drees, T. Kornberg, J. M. Bishop, *Cell* **42**, 831 (1985).
90. F. M. Hoffman, L. D. Fresco, H. Hoffman-Falk, B.-Z. Shilo, *ibid.* **35**, 393 (1983).
91. R. J. Gregory, K. L. Kammermeyer, W. S. Vincent III, S. G. Wadsworth, *Mol. Cell. Biol.* **7**, 2119 (1987).
92. E. Shtivelman, B. Lifshitz, R. P. Gale, B. A. Roe, E. Canaani, *Cell* **47**, 277 (1986).
93. G. D. Kruh *et al.*, *Science* **234**, 1545 (1986).
94. J. M. Goddard, J. J. Weiland, M. R. Capecchi, *Proc. Natl. Acad. Sci. U.S.A.* **83**, 2172 (1986).
95. A. J. M. Roebroek *et al.*, *EMBO J.* **4**, 2897 (1985).
96. A. J. M. Roebroek, J. A. Schalken, C. Onnekink, H. P. J. Bloemers, W. J. M. Van de Ven, *J. Virol.* **61**, 2009 (1987).
97. C.-C. Huang, C. Hammond, J. M. Bishop, *J. Mol. Biol.* **181**, 175 (1985).
98. A. Ullrich *et al.*, *Nature* **309**, 418 (1984).
99. L. Coussens *et al.*, *Science* **230**, 1132 (1985); T. Yamamoto *et al.*, *Nature* **319**, 230 (1986).
100. C. I. Bargmann, M.-C. Hung, R. A. Weinberg, *Nature* **319**, 226 (1986).
101. E. Livneh, L. Glazer, D. Segal, J. Schlessinger, B.-Z. Shilo, *Cell* **40**, 599 (1985).
102. A. Ullrich *et al.*, *Nature* **313**, 756 (1985); Y. Ebina *et al.*, *Cell* **40**, 747 (1985).
103. A. Ullrich *et al.*, *EMBO J.* **5**, 2503 (1986).
104. Y. Nishida, M. Hata, Y. Nishizuka, W. J. Rutter, Y. Ebina, *Biochem. Biophys. Res. Commun.* **141**, 474 (1986).
105. H. Matsushime, L.-H. Wang, M. Shibuya, *Mol. Cell. Biol.* **6**, 3000 (1986); C. Birchmeier, D. Birnbaum, G. Waitches, O. Fasano, M. Wigler, *ibid.*, p. 3109.
106. W. S. Neckameyer, M. Shibuya, M.-T. Hsu, L.-H. Wang, *ibid.*, p. 1478.
107. S. B. Podell and B. M. Sefton, *Oncogene* **2**, 9 (1987).
108. E. Hafen, K. Basler, J.-E. Edstroem, G. M. Rubin, *Science* **236**, 55 (1987).
109. D. Martin-Zanca, S. H. Hughes, M. Barbacid, *Nature* **319**, 743 (1986).
110. M. Park *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **84**, 6379 (1987); A. M.-L. Chan *et al.*, *Oncogene* **1**, 229 (1987).
111. Y. Yarden *et al.*, *Nature* **323**, 226 (1986).
112. L. Coussens *et al.*, *ibid.* **320**, 277 (1986).
113. Y. Yarden *et al.*, *EMBO J.* **6**, 3341 (1987).
114. M. Takahashi and G. M. Cooper, *Mol. Cell. Biol.* **7**, 1378 (1987).
115. D. A. Foster, J. B. Levy, G. Q. Daley, M. C. Simon, H. Hanafusa, *ibid.* **6**, 325 (1986).
116. D.-F. Feng, M. S. Johnson, R. F. Doolittle, *J. Mol. Evol.* **21**, 112 (1985).
117. W. M. Fitch and E. Margoliash, *Science* **155**, 279 (1967).
118. D.-F. Feng and R. F. Doolittle, *J. Mol. Evol.* **25**, 351 (1987).
119. We thank D.-F. Feng and R. F. Doolittle for much helpful advice concerning phylogenetic tree construction, R. W. Holley for support, J. R. Woodgett for making unpublished sequence data available, R. A. Lindberg and L. T. Williams for allowing us to reference their unpublished work, K. Hyde for assistance in sequence data entry, and L. Norris for help in preparation of Fig. 1. Supported by grant GM38793 from the NIH (SKH).

Exhibit 9

Biologically Active Mutants with Deletions in the v-*mos* Oncogene Assayed with Retroviral Vectors

RICHARD J. BOLD AND DANIEL J. DONOGHUE*

Department of Chemistry, University of California at San Diego, La Jolla, California 92093

We have constructed retroviral expression vectors by manipulation of the Moloney murine leukemia virus genome such that an exogenous DNA sequence may be inserted and subsequently expressed when introduced into mammalian cells. A series of N-terminal deletions of the v-*mos* oncogene was constructed and assayed for biological activity with these retroviral expression vectors. The results of the deletion analysis demonstrate that the region of p37^{mos} coding region upstream of the third methionine codon is dispensable with respect to transformation. However, deletion mutants of v-*mos* which allow initiation of translation at the fourth methionine codon have lost the biological activity of the parental v-*mos* gene. Furthermore, experiments were also carried out to define the C-terminal limit of the active region of p37^{mos} by the construction of premature termination mutants by the insertion of a termination oligonucleotide. Insertion of the oligonucleotide just 69 base pairs upstream from the wild-type termination site abolished the focus-forming ability of v-*mos*. Thus, we have shown the N-terminal limit of the active region of p37^{mos} to be between the third and fourth methionines, while the C-terminal limit is within the last 23 amino acids of the protein.

Moloney murine sarcoma virus (M-MSV) was originally isolated from a sarcoma which appeared after injection of Moloney murine leukemia virus (M-MLV) into BALB/c mice (32). M-MSV arose by recombination between nondefective M-MLV and normal mouse cellular DNA (4, 13–15). The acquisition of the cellular DNA occurred concomitantly with significant deletions of the parental M-MLV genome. Thus, M-MSV is able to transform fibroblasts but is defective for viral replication (1, 4, 35–37).

The acquired mouse cellular DNA of M-MSV has been mapped to an uninterrupted sequence of approximately 1,200 base pairs (bp) near the 3' terminus of the M-MSV genome (14, 23, 47) and is referred to as v-*mos*. The nucleotide sequence of v-*mos* reveals that the 1,125-bp open reading frame is an *env-mos* fusion such that the first five codons of v-*mos* are derived from the M-MLV *env* gene and the remainder are derived from c-*mos* (11, 45, 46).

A 37,000-dalton protein was first identified as the v-*mos* gene product by in vitro transcription of M-MSV viral RNA (34). Subsequently, a 37,000-dalton protein was immunoprecipitated from M-MSV-transformed cells with an antibody directed against a peptide corresponding to the predicted C terminus of the v-*mos* gene product (36, 37). This protein, referred to as p37^{mos}, is presumably responsible for transformation by M-MSV. p37^{mos} exhibits limited regions of amino acid sequence homology with the catalytic subunit of cyclic AMP (cAMP)-dependent protein kinase (3). While other oncogenic proteins which show a similar homology with the cAMP-dependent protein kinase possess protein kinase activity, such as p60^{src}, no enzymatic activity has been unequivocally demonstrated for p37^{mos} (27).

We have constructed a series of deletions in the N-terminal coding region of v-*mos*. In addition, premature termination mutants of the v-*mos* gene were constructed. These mutations define the region of the gene requisite for transformation. The mutants of v-*mos* were expressed in eucaryotic cells with retroviral expression vectors. The retroviral vectors described here represent deleted deriva-

tives of M-MLV constructed such that the inserted DNA fragment replaces the retroviral *env* gene. Our work allows us to define the limits of the v-*mos* gene which are required to encode a transforming gene product. Mutants of v-*mos* with N-terminal deletions which allow initiation of translation at the third in-frame ATG of v-*mos* still retain biological activity, whereas mutants with more extensive N-terminal deletions are biologically inactive. Premature termination mutants have demonstrated that some portion of the C-terminal 69 nucleotides is necessary to maintain biological activity.

MATERIALS AND METHODS

Construction of retroviral expression vectors. Many of the details of the retroviral vectors are summarized in Fig. 1. The proviral clone of M-MLV, p836, was initially described by Hoffman et al. (22) and served as the parental plasmid for the vectors described in this work. We desired a unique *Xho*I restriction site in our vectors, which was achieved by the following steps. First, the *Xho*I-*Hind*III fragment (nucleotides [nt] 1560 to 4894 in the M-MLV sequence, reference 39), which encompasses the *gag-pol* region of M-MLV, was deleted. This was accomplished by blunt-end ligation of the *Xho*I terminus to the *Hind*III terminus after treatment with the Klenow fragment of DNA polymerase I. This protocol destroyed the *Xho*I site but restored the *Hind*III site. Second, the *Xho*I site downstream of the 3' long terminal repeat (LTR) was removed by digestion with *Xho*I, followed by treatment with Klenow fragment, and religation. Third, the *Cla*I site in the *env* gene (nt 7674) was converted to an *Xho*I site by the insertion of an *Xho*I linker (CCTCGAGG). It should be noted that there is also a second *Cla*I site in M-MLV, at nt 4980, but this site is methylated in DNA grown in *dam$^+$* strains of *Escherichia coli* and is refractory to cleavage by *Cla*I.

To position the *Xho*I restriction site at the correct location in the vectors, the following sequence of steps was undertaken. First, the *Hind*III-*Hha*I fragment of M-MLV (nt 4894 to 5780 of M-MLV), which contains the 3' splice acceptor site of the *env* gene and also the initiator ATG for the *env* gene product, was subcloned as an *Hind*III-*Eco*RI fragment

---

* Corresponding author.

to yield the plasmid pDD90. Furthermore, the *Eco*RI site was converted to an *Xho*I site by linker insertion (CCTCGAGG) to yield the clone pDD92. Second, the subclone pDD90 was linearized at the *Eco*RI site, which
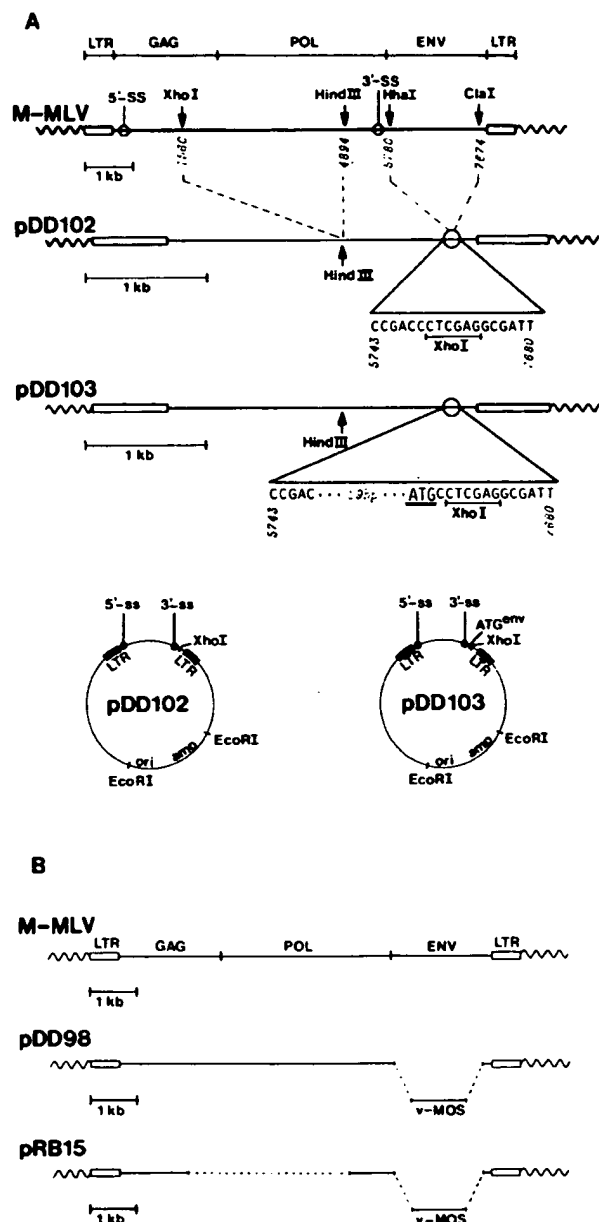

B

FIG. 1. (A) Construction of retroviral expression vectors. Shown is the genome of M-MLV with the splice sites and the retroviral genes denoted. Restriction sites used in the construction of the expression vectors are also shown as well as the corresponding nucleotide positions (39). Details of the construction of pDD102 and pDD103 are given in Materials and Methods. Each vector contains a unique *Xho*I site for insertion of a DNA fragment in place of the retroviral *env* gene. In the vector pDD102, the *env* gene initiation codon has been removed, so a DNA insert must provide its own ATG codon for the initiation of translation. In the vector pDD103, the *env* gene initiation codon remains and will be fused to

replaced the former *Hha*I site (at nt 5780), and treated with BAL 31 nuclease (28). After BAL 31 digestion, *Xho*I linkers (CCTCGAGG) were ligated onto the digested DNA. Third, several of the resulting plasmids were isolated, and the position of the *Xho*I linker was determined in each by nucleotide sequencing (30). One such clone, pDD94, had 29 bp removed by BAL 31, including the *env* gene ATG. In this clone, the *Xho*I site was located at nt 5751 of M-MLV, whereas the *env* gene ATG was located at nt 5777. The final step in the construction of the vectors was to replace the *Hind*III-*Xho*I fragment of the M-MLV derivative pDD99 with the *Hind*III-*Xho*I fragment of each of the subclones pDD94 and pDD92 described above. Insertion of the *Hind*III-*Xho*I fragment of pDD94 gave rise to the vector pDD102, while insertion of the *Hind*III-*Xho*I fragment of pDD92 gave rise to the vector pDD103.

The vectors pDD102 and pDD103 are similar in the following details. (i) The *Xho*I-*Hind*III region of M-MLV (nt 1560 to 4894) is deleted. (ii) Most of the *env* gene, from *Hha*I to *Cla*I (nt 5780 to 7674), is deleted. (iii) The 5' splice donor site and the 3' splice acceptor site for the *env* gene are retained. The two vectors differ in that pDD102 lacks the *env* gene ATG codon, whereas pDD103 retains the *env* gene ATG codon.

The clone pDD98 is similar to the clone pRB15, shown in Fig. 2, which contains the wild-type v-*mos* gene inserted into the vector pDD102. pDD98 differs from pRB15, however, by the retention of the *gag-pol* region of M-MLV, contained on the *Xho*I-*Hind*III fragment (nt 1560 to 4894 in the M-MLV sequence, reference 39), as shown in Fig. 1B.

Construction of deletion mutants in the *mos* gene. The v-*mos* gene which served as the starting material for the isolation of deletion mutants was the *Xba*I to *Hind*III fragment of the 124-MSV strain contained in the plasmid pDD0 (11). The long open reading frame of the v-*mos* gene which encodes p37^{mos} is conveniently flanked by a unique *Xba*I site upstream and a unique *Hind*III site downstream. The *Hind*III restriction site downstream of the coding region was first converted to an *Xho*I site by insertion of an *Xho*I linker (CCTCGAGG). The resulting plasmid was then linearized at the *Xba*I site upstream of the coding region and treated with BAL 31 exonuclease (28) for varying lengths of time. *Xho*I linkers were then ligated to the DNA, yielding *mos* genes deleted for various distances and flanked by *Xho*I sites. The exact length of deletion was determined by Maxam-Gilbert sequencing (30) with 3'-end-labeled restriction fragments. The endpoint of each deletion, as determined by sequencing, is indicated in Fig. 2.

Construction of C-terminal premature termination mutants. The v-*mos* gene was linearized at one of three unique restriction sites, *Kpn*I, *Hpa*I, or *Sst*II, in the long open reading frame (see Fig. 3). The linear fragments were then

any DNA fragment inserted at the *Xho*I site. The lower portion of the figure shows the entire plasmids with the splice sites, LTR sequences, and important restriction sites denoted. Both pDD102 and pDD103 are approximately 15.8 kilobases in size. (B) Comparison of two vectors which express v-*mos*. The genome structures of pDD98 and pRB15 are compared with the parental retroviral genome of M-MLV. pDD98 retains the complete *gag-pol* region of M-MLV, whereas most of the *gag-pol* region has been deleted from pRB15. Given that pRB15 demonstrated greater biological activity than pDD98 (see text and Table 1), a vector with this same deletion (pDD102) was used for the assay of all v-*mos* deletion mutants.

treated with the Klenow fragment of DNA polymerase in the presence of all four deoxyribonucleoside triphosphates to yield blunt ends. The termination oligonucleotide (TCAATCAGTCAAGCTTGACTGATTGA) was then ligated onto the DNA. This oligonucleotide is self-complementary and contains termination codons in all three reading frames. The presence of the oligonucleotide in the resulting plasmids was monitored by the acquisition of an HindIII site contained within the oligonucleotide.

Biological assay. The biological activity of the various mutants of the v-mos gene was assayed with the expression vector pDD102. The XhoI-XhoI fragment of each mutated gene was inserted into the unique XhoI site of pDD102. Colony hybridization was used to ensure the presence of an insert, and restriction mapping was used to define the correct orientation of the insert (20).
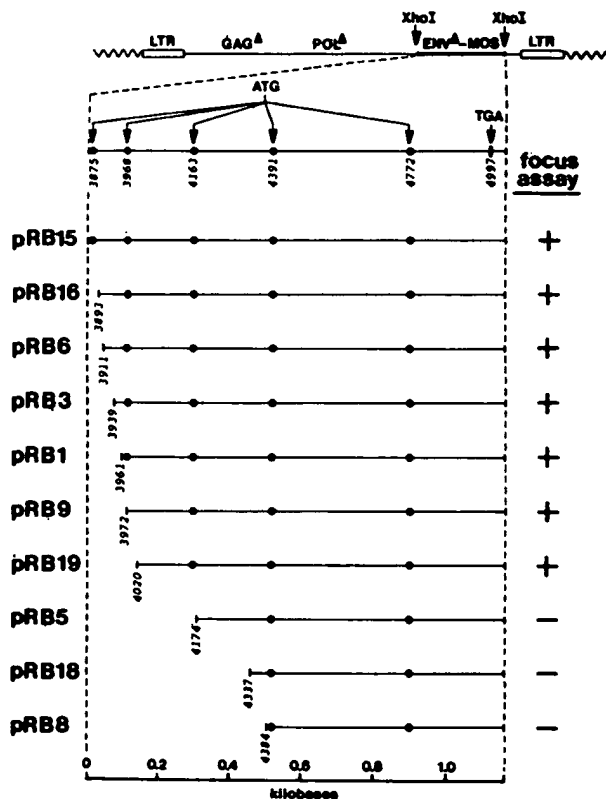
TABLE 1. Transforming activities of deletion mutants[a]

Plasmid	First ATG remaining	Site of termination linker	Biological activity (FFU/pmol)	Rescued virus titer (FFU/ml)
pDD98	1		7.4×10^2	1.7×10^3
pRB15 (wild type)	1		1.3×10^4	3.2×10^5
pRB16	2		1.2×10^3	6.0×10^4
pRB6	2		1.1×10^3	9.5×10^4
pRB3	2		1.0×10^3	9.0×10^4
pRB1	2		1.0×10^3	9.7×10^4
pRB9	3		1.1×10^3	1.1×10^5
pRB19	3		1.1×10^3	1.2×10^5
pRB5	4		$<6.3 \times 10^1$	$<5.0 \times 10^0$
pRB18	4		$<8.3 \times 10^1$	$<5.0 \times 10^0$
pRB8	4		$<6.0 \times 10^1$	$<5.0 \times 10^0$
pRB27		KpnI	9.7×10^1	4.0×10^1
pRB28		HpaI	1.3×10^1	$<1.0 \times 10^1$
pRB26		SstII	4.1×10^1	$<2.5 \times 10^1$

[a] The upper portion of the table indicates the biological activity of the N-terminal deletion mutants. The table also denotes the first in-frame ATG codon available for initiating translation in each of the deletion mutants. The biological activity of each plasmid was assayed as described in Materials and Methods. Values shown are calculated as focus-forming units per picomole (FFU/pmol) of transfected DNA. The lower portion of the table shows the biological activity of the premature termination mutants. The restriction sites at which the termination linker was inserted are denoted. The virus titer of collected culture fluid from the transfection plates was determined as described in Materials and Methods. The titers shown are calculated as focus-forming units per milliliter of culture fluid (FFU/ml) which was collected approximately 7 days after transfection.

FIG. 2. N-terminal deletion analysis of the v-mos gene. The structure of the M-MLV-derived expression vector pDD102 is shown with the full-length wild-type v-mos gene inserted as an XhoI-XhoI fragment. The long open reading frame of v-mos has been expanded to show the five ATG codons in the 1125-bp coding sequence. The ATG codons are represented by solid circles, and the opal terminator codon is also shown. Nucleotide numbers associated with the ATG codons and the terminator codon are from the published sequence of 124-MSV (46). The nucleotide position at each deletion endpoint is indicated, as determined by Maxam-Gilbert sequencing; the numbers refer to the first remaining base pair of v-mos after the XhoI linker. The biological activity of each deletion mutant was assayed by its ability to induce foci when introduced into NIH 3T3 cells. A plus indicates a minimum value of 10^3 FFU/pmol of cloned DNA. A minus indicates a maximum value of 8.3×10^1 FFU/pmol.

The general protocol used for the focus assay was that described by Graham and Van der Eb (19). A 1-μg portion of each DNA, together with 0.5 μg of DNA of the replication-competent clone of M-MLV, p836 (22), was transfected in the presence of 15 μg of sheared calf thymus DNA. Each sample was applied to a 5-cm plate of approximately 50% confluent NIH 3T3 cells as a calcium phosphate coprecipitate. After approximately 12 to 18 h, the plates were trypsinized, and the cells were distributed to four 10-cm plates. The focus assays were then scored 12 to 14 days after transfection. Results of the focus assays are shown in Table 1. These results were also confirmed by collecting the culture fluid from the transfection plates and determining titers for this culture fluid on fresh monolayers of NIH 3T3 cells for focus-forming virus. The results of the transforming virus titers are shown in Table 1 and were always consistent with the results of the primary transfection assays.

RESULTS

Construction of retroviral expression vectors. We constructed retroviral vectors which permit the expression in eucaryotic cells of an inserted gene (2, 8, 10, 17, 18, 21, 24, 33, 43). The vectors are constructed in such a way that the inserted DNA fragment is substituted for the env gene of M-MLV (21, 39). Consequently, the mRNA of the inserted gene will be transcribed and spliced in a fashion identical to that of the M-MLV env gene. Furthermore, the vectors provide LTRs both 5' and 3' to the insert which are required for the enhancement of transforming activity of v-mos as well as recovery of infectious virus (4, 44, 49). The two vectors described here, pDD102 and pDD103, contain 5' and 3' LTRs, 5' and 3' splice sites for the env gene, similar deletions, and a unique XhoI site for insertion of a DNA fragment (Fig. 1). However, pDD102 and pDD103 differ by the absence or presence of the env gene ATG codon.

In the vector pDD103, a unique XhoI site was inserted

directly downstream of the *env* gene ATG codon so that an inserted gene need not provide its own ATG for initiation of translation when expressed. In contrast, pDD102 lacks the *env* gene initiation codon. Treatment with BAL 31 nuclease was used to remove the *env* gene ATG and a short region upstream (Fig. 1). The expression of a gene inserted at the unique *Xho*I site of pDD102 requires that the inserted gene provide its own initiation codon for translation.

The parental v-*mos* gene was inserted into the vector pDD102 to yield the plasmid pRB15. This plasmid was biologically active when transfected into NIH 3T3 cells and scored in a conventional focus assay (Table 1). When the transfection was carried out in the presence of DNA of a clone of infectious M-MLV, it was possible to demonstrate transmissible transforming virus in the culture fluid from the transfected cells. We also constructed another retroviral vector, pDD98, which contains the wild-type v-*mos* gene. This construct is similar to pRB15 but retains the complete *gag-pol* region from M-MLV. The genome structures of M-MLV, pRB15, and pDD98 are summarized in Fig. 1B. pDD98 consistently yielded fewer foci (7.4 × 10² foci per pmol) compared with pRB15 (1.3 × 10⁴ foci per pmol) (Table 1). Also, the titer of recovered transforming virus was lower for pDD98 (1.7 × 10³ focus-forming units [FFU]/ml) compared with pRB15 (3.2 × 10⁵ FFU/ml). The low transformation frequency of pDD98 was shown not to be due to the reported cytotoxicity of v-*mos* (37), as the transformation efficiency of the herpesvirus thymidine kinase gene was not lowered by cotransfection with pDD98 (data not shown). Due to the higher biological activity of the clone pRB15 compared with the clone pDD98, we decided to use the vector pDD102 for subsequent analysis of the v-*mos* deletion mutants described below.

N-terminal deletions of v-*mos*. To delimit the region of the v-*mos* gene necessary for neoplastic transformation, we constructed a series of deletions of the N-terminal region. The N-terminal region of v-*mos* was treated with BAL 31 nuclease for varied amounts of time to yield deletions in this region. The position of the deletion endpoints varied from just beyond the first ATG (pRB16), a deletion of 18 bp, to just before the fourth ATG (pRB8), which has 510 bp of v-*mos* removed (Fig. 2). The deleted v-*mos* genes, after insertion into the expression vector pDD102, were introduced into NIH 3T3 cells by the calcium phosphate coprecipitation technique to test the effect of each deletion on the transforming activity (19).

The biological activity of these deletions, as determined by standard focus assay, is indicated by + or − in Fig. 2, and quantitation is presented in Table 1. Deletion mutants with transforming activities denoted as − did not possess any transforming activity above the level of detection. This was confirmed by assaying the culture fluids for the presence of transmissible focus-forming virus, as described in Table 1. Mutants with deletions of only the first ATG codon of the open reading frame, e.g., pRB1, as well as mutants with deletions of both the first and second ATG codons, e.g., pRB9, were still biologically active. However, deletions extending beyond the third ATG codon abolished biological activity. Therefore, v-*mos* deletion clones such as pRB9 and pRB19, which presumably initiate translation at the third ATG, produce polypeptides capable of transforming NIH 3T3 cells. This is in contrast to the v-*mos* deletion clones, such as pRB8, which must initiate translation at the fourth ATG and are incapable of transformation. Thus, the region of v-*mos* upstream of the third ATG codon is unnecessary for neoplastic transformation activity.

It must be noted that the transforming ability of the deletions is below the level of wild-type v-*mos* activity. As observed in Table 1, there is approximately a 10-fold reduction in the transforming activity of the v-*mos* deletions below that of the wild-type gene. This might result from initiation at out-of-frame ATG codons which are upstream from the correct, in-frame ATG (11). However, this seems unlikely as all the transforming deletions exhibit the same reduction in transforming activity, whereas they do not all possess out-of-frame ATG codons upstream of the first remaining in-frame initiation codon. Another possible reason for the decrease in activity involves protein structure. While the N-terminal region upstream of the third ATG may not be absolutely required for transforming activity, it may play a significant role in the stability or the correct tertiary structure of the protein. The important point to note, though, is that the mutants with deletions up to but not including the third ATG codon still possess the ability to transform cells, albeit not as efficiently as the wild-type parent.

A cytotoxic effect has been reported during acute infection of NIH 3T3 cells by 124-MSV (37). However, periodic examination of the transfected cells by phase-contrast microscopy revealed no evidence of extensive cell death, as would be expected if any of these constructs resulted in cytotoxicity.

Premature termination of v-*mos* by oligonucleotide insertion. We also constructed mutations of the v-*mos* gene specifying premature termination of translation. A self-complementary oligonucleotide (Fig. 3) was synthesized which contains opal termination codons in all three reading frames. This oligonucleotide was then inserted at three
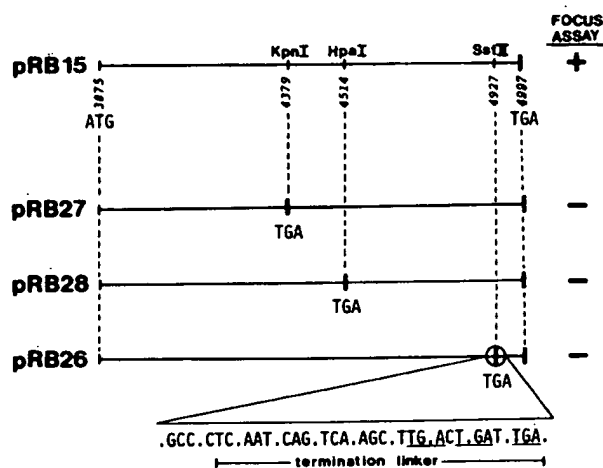

FIG. 3. Premature termination of the v-*mos* gene by oligonucleotide insertion. The full-length coding region of the v-*mos* gene is shown from the initiation codon (ATG) to the termination codon (TGA). The nucleotide positions of these codons and of the restriction sites indicated refer to the sequence of 124-MSV (46). The synthetic oligonucleotide as shown was inserted at each of three restriction sites so as to cause premature termination. Upon insertion of this linker at any restriction site, regardless of reading frame, translation will be terminated due to the presence of termination codons in all three possible reading frames. The biological activity of the premature termination mutants was assayed by their ability to induce focus formation in NIH 3T3 cells. All termination mutants possessed no biological activity and had a maximum value of 9.7 × 10¹ FFU/pmol.

unique restriction sites in v-*mos* (Fig. 3). These mutants were assayed for focus-forming activity in the vector pDD102, and all were biologically inactive. Thus, the C-terminal limit of the gene required for the transforming ability of v-*mos* is beyond the *Sst*II site at the 352nd amino acid of the protein, since termination at this site destroys biological activity. This indicates that some portion of the 23 amino acids of the C terminus is essential for v-*mos* to transform NIH 3T3 cells.

DISCUSSION

Retroviral expression vectors have become the method of choice for the introduction of foreign genetic material into eucaryotic cells (8, 10, 21, 24, 26, 31, 33, 38, 41, 43, 48). The major characteristics which can be used to differentiate the various retroviral expression vectors are: (i) the retroviral gene which is replaced with the exogenous gene, (ii) the presence or absence of a selectable marker within the expression vector, and (iii) whether the expression vector provides a promoter for the inserted gene.

Our expression vectors, derived by deletion of M-MLV, allow for the replacement of the viral *env* gene sequence by the desired foreign gene while maintaining both LTR sequences, splicing signals, and packaging signals. The expression vectors are replication defective and therefore must be accompanied by coinfection by a helper virus if a transmissible virus stock is desired (29). While the vectors we constructed do not have a selectable marker other than the gene insert, there exists a unique *Hin*dIII restriction site in the deleted *gag-pol* region for the insertion of such a selection marker. Two other groups have constructed retroviral expression vectors which are similar to those described in this work and deserve special note. Cepko et al. (8) have constructed an expression vector derived from M-MLV such that a foreign gene may be inserted so as to replace the retroviral *env* gene. A special feature of this vector is the presence of the Tn5 transposon-derived neomycin resistance gene which replaces the *gag-pol* region of the parental retrovirus. Hwang and Gilboa (24) constructed M-MLV-derived vectors which replace the *env* gene of the retrovirus with the neomycin resistance gene derived from Tn5. Our vectors, however, contain deletions within the *gag-pol* region, whereas the vectors of Hwang and Gilboa retain intact the *gag-pol* region.

In this work, we described two very similar retroviral constructs, pDD98 and pRB15, which express the v-*mos* oncogene but which differ by the presence or absence of the *gag-pol* region. Hwang et al. (25) have previously reported the existence of two regions in the *gag-pol* region which are required for efficient *env* gene expression. However, when the v-*mos* oncogene is inserted into the deleted vector, expression is more efficient than when inserted in the nondeleted vector. This increased expression in the deleted vector is reflected both in higher numbers of foci per picomole of transfected DNA and higher titers of transmissible transforming virus which is recovered from transfected cells (Table 1). These results suggest that the presence of the *gag-pol* intron is not essential, when using the retroviral vectors described here, for efficient expression of genes substituted in place of the retroviral *env* gene.

Through the use of retroviral expression vectors which we have constructed, we have been able to assay the biological activity of N-terminal deletions of v-*mos*. The results of this work show that some region between the third and fourth initiation codons is absolutely required for induction of focus formation by v-*mos*. Thus, the remainder of the gene down-

stream from the third ATG codon possesses all regions requisite for biological activity. This defines the N-terminal limit of the v-*mos* gene necessary for focus formation.

Previous delimitation of the v-*mos* gene required for biological activity has relied on altered recombinational events of the *mos* gene (6, 13, 14, 16, 40, 42). Myeloproliferative sarcoma virus is a member of the MSV family originating upon serial transplantation of a tumor induced in a newborn mouse by uncloned MSV with M-MLV as a helper (42). In the myeloproliferative sarcoma virus genome, a frameshift mutation in the N-terminal coding region produces a truncated gene product of 19 amino acids upon initiation of translation at the first ATG of the v-*mos* gene. However, in myeloproliferative sarcoma virus, initiation can begin at the second ATG of *mos* and continue through the proper C terminus, presumably generating the gene product responsible for transformation. This work with myeloproliferative sarcoma virus places the active region of the v-*mos* gene beyond the second ATG.

Another altered recombinational event involving the *mos* gene was discovered by Canaani et al. when a rearranged c-*mos* gene (rc-*mos*) was isolated from a mouse myeloma tumor (7, 9, 40). The rearrangement involved the deletion of 263 nucleotides from the 5' coding sequence of c-*mos* and the substitution of this information with a new cellular DNA sequence. Thus the gene product from rc-*mos* has an N terminus of 28 amino acids donated by the substituted cellular sequence while the remainder of the protein is from the c-*mos* gene. The point of recombination is at the 73rd amino acid of v-*mos*, which lies between the second and third ATG codons of the 124-MSV v-*mos* gene. This work demonstrates that the transforming potential of c-*mos* lies downstream of the 73rd amino acid of the full-length gene product (7, 9, 40). Our work is consistent with these earlier observations by placing the N-terminal limit of the biologically active region of the v-*mos* gene product at the third ATG, which codes for the 97th amino acid of the v-*mos* protein.

Using a termination oligonucleotide, we have shown that some portion of the 23 amino acids at the C terminus of p37^{mos} is necessary for biological activity. The sites of recombination among different variants of M-MSV have provided the only previous delimitation of the C-terminal active region of v-*mos* (13, 14). The m1-MSV strain has a small internal deletion at the C terminus which results in an altered amino acid sequence beyond the site of deletion (6). Figure 4 shows the point of divergence between m1-MSV and 124-MSV. However, of the C-terminal 23 amino acids which we have shown to contain a region necessary for transformation, 17 are conserved in m1-MSV. Since only those variants of MSV which still possess the transforming capabilities of v-*mos* will be isolated, the recombinational site at the C terminus can be used to delimit the region necessary for biological activity (6, 13, 14). Our work, as well as the absence of MSV variants with C-terminal recombinational sites upstream of these terminal 23 amino acids, demonstrates that some portion of this sequence is required for biological activity of v-*mos*.

The v-*mos* gene product is one of several oncogenic proteins which show homology to the bovine cAMP-dependent protein kinase but possess no specific tyrosine kinase activity (3, 27). There are several specific regions within the v-*mos* gene product which show conservation of residues with the protein kinase (Fig. 4). Directly downstream of the third ATG codon is a sequence of nine amino acids, of which six are conserved with the cAMP-dependent
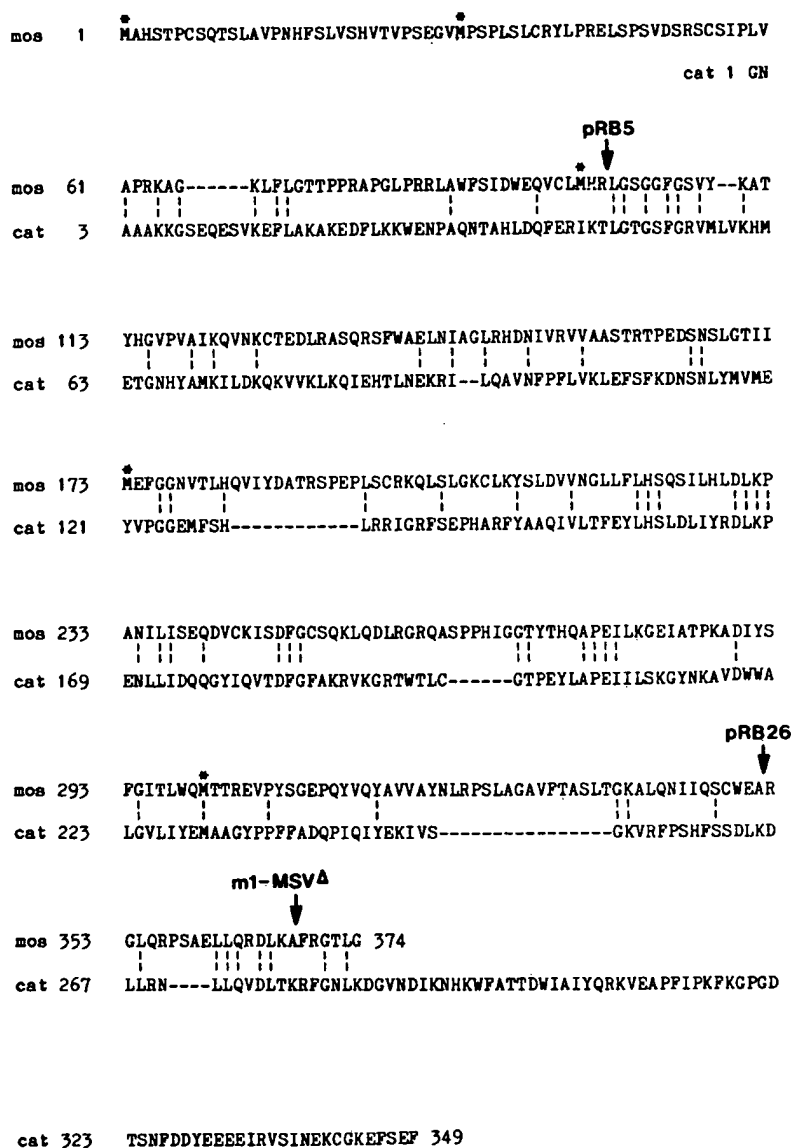
```
mos    1    MAHSTPCSQTSLAVPNHFSLVSHVTVPSEGVMPSPLSLCRYLPRELSPSVDSRSCSIPLV

                                                        cat 1 GN


                                                  pRB5
                                                    ↓
mos   61    APRKAG------KLFLGTTPPRAPGLPRRLAWPSIDWEQVCLMHRLGSGGFGSVY--KAT
            I  I  I       I  II               I             I    II I II I I   I
cat    3    AAAKKGSEQESVKEPLAKAKEDPLKKWENPAQNTAHLDQFERIKTLGTGSFGRVMLVKHM



mos  113    YHGVPVAIKQVNKCTEDLRASQRSFWAELNIAGLRHDNIVRVVAASTRTPEDSNSLGTII
            I  I I  I    I               I  I  I  I   I    I             II
cat   63    ETGNHYAMKILDKQKVVKLKQIEHTLNEKRI--LQAVNFPPLVKLEFSFKDNSNLYMVME



mos  173    MEFGGNVTLHQVIYDATRSPEPLSCRKQLSLGKCLKYSLDVVNGLLFLHSQSILHLDLKP
            II  I         I    I   II    I    I    I  III       IIII
cat  121    YVPGGEMFSH------------LRRIGRFSEPHARFYAAQIVLTFEYLHSLDLIYRDLKP



mos  233    ANILISEQDVCKISDPGCSQKLQDLRGRQASPPHIGGTYTHQAPEILKGEIATPKADIYS
            I III I   III                 II   IIII          I
cat  169    ENLLIDQQGYIQVTDFGPAKRVKGRTWTLC------GTPEYLAPEIILSKGYNKAVDWWA


                                                        pRB26
                                                          ↓
mos  293    FGITLWQMTTREVPYSGEPQYVQYAVVAYNLRPSLAGAVPTASLTGKALQNIIQSCWEAR
            I          I        I              II            I
cat  223    LGVLIYEMAAGYPPFPADQPIQIYEKIVS---------------GKVRFPSHFSSDLKD


             m1-MSVΔ
                ↓
mos  353    GLQRPSAELLQRDLKAFRGTLG 374
            I        III  II    I  I
cat  267    LLRN----LLQVDLTKRFGNLKDGVNDIKNHKWFATTDWIAIYQRKVEAPFIPKFKGPGD



cat  323    TSNFDDYEEEEIRVSINEKCGKEFSEF 349
```

FIG. 4. Location of v-*mos* mutants and relationship to homology with cAMP-dependent protein kinase. The amino acid sequence of p37^{mos} is shown (in one-letter code) in alignment with the catalytic subunit of cAMP-dependent protein kinase (cat) (3). Asterisks above the *mos* sequence denote the five methionine residues within the protein. The endpoint of deletion mutant pRB5 represents the smallest deletion which destroys biological activity. The site of termination linker insertion in pRB26 is also shown. This leads to truncation of 23 amino acids from the C terminus of p37^{mos} and the loss of biological activity. The point of divergence between 124-MSV and m1-MSV is also shown. The closely related strain m1-MSV encodes a gene product with a slightly altered C terminus, due to a shift in the reading frame near the C terminus. As a result, the gene product of m1-MSV retains only the first 17 residues of the 23 C-terminal residues of the 124-MSV gene product.

protein kinase (5). Figure 4 shows the N-terminal endpoint of the smallest deletion mutant, pRB5, which is incapable of transformation. This mutant initiates translation at the fourth in-frame ATG when expressed in the vector pDD102.

It should be noted that a presumptive ATP-binding site in p37^{mos} lies within the region which is required for transformation, as shown by the deletion analysis presented here. This presumptive ATP-binding site, which includes lysine residue 121 of p37^{mos}, occurs just downstream of the third methionine. A similar ATP-binding site occurs in the catalytic subunit of the cAMP-dependent protein kinase and also in p60^{src}, and site-directed mutagenesis of lysine 121 of

p37^{mos} indicates that this residue is functionally important for transformation by v-*mos* (M. Hannink and Daniel J. Donoghue, manuscript submitted for publication).

Figure 4 also shows the site at which the biologically inactive C-terminal mutant pRB26 terminates translation because of the termination linker insertion. Beyond the site of premature termination, another region of v-*mos* is highly conserved with the protein kinase in which five of six residues are identical (reference 3; Fig. 4). This sequence homology lies within the C-terminal region which is required for transformation by v-*mos*. Although the v-*mos* gene product function has not been identified, it appears to

require regions which have conserved amino acid sequences with cAMP-dependent protein kinase.

The normally quiescent c-*mos* gene, when acquired and expressed by the M-MSV retrovirus, becomes acutely transforming (5, 7, 9, 16). Since the function of the v-*mos* gene product has yet to be determined, neither its original role as a c-*mos* gene product nor its method of transformation when expressed by a retrovirus is known (12, 27, 35). Our work delineates the N- and C-terminal regions of the *mos* gene requisite for biological activity and consequently the sequences which are indispensable for an active v-*mos* gene product.

ACKNOWLEDGMENTS

LITERATURE CITED

1. Andersson, P., M. P. Goldfarb, and R. A. Weinberg. 1979. A defined subgenomic fragment of in vitro synthesized Moloney sarcoma virus DNA can induce cell transformation upon transfection. Cell 16:63–75.

2. Bandyopadhyay, P. K., and H. M. Temin. 1984. Expression of complete chicken thymidine kinase gene inserted in a retrovirus vector. Mol. Cell. Biol. 4:749–754.

3. Barker, W. C., and M. O. Dayhoff. 1982. Viral *src* gene products are related to the catalytic chain of mammalian cAMP-dependent protein kinase. Proc. Natl. Acad. Sci. USA 79:2836–2839.

4. Blair, D. G., W. L. McClements, M. K. Oskarsson, P. J. Fischinger, and G. F. Vande Woude. 1980. Biological activity of cloned Moloney sarcoma virus DNA: terminally redundant sequences may enhance transformation efficiency. Proc. Natl. Acad. Sci. USA 77:3504–3508.

5. Blair, D. G., M. Oskarsson, T. G. Wood, W. L. McClements, P. J. Fischinger, and G. F. Vande Woude. 1981. Activation of the transforming potential of a normal cell sequence: a molecular model for oncogenesis. Science 212:941–943.

6. Brow, M. A. D., A. Sen, and J. G. Sutcliffe. 1984. Nucleotide sequence of the transforming gene of m1 murine sarcoma virus. J. Virol.49:579–582.

7. Canaani, E., O. Dreazen, A. Klar, G. Rechavi, D. Ram, J. B. Cohen, and D. Givol. 1983. Activation of the c-*mos* oncogene in a mouse plasmacytoma by insertion of an endogenous intracisternal A-particle genome. Proc. Natl. Acad. Sci. USA 80:7118–7122.

8. Cepko, C. L., B. E. Roberts, and R. C. Mulligan. 1984. Construction and applications of a highly transmissible murine retrovirus shuttle vector. Cell 37:1053–1062.

9. Cohen, J. B., T. Unger, G. Rechavi, E. Canaani, and D. Givol. 1983. Rearrangement of the oncogene c-*mos* in mouse myeloma NSI and hybridomas. Nature (London) 306:797–799.

10. Cone, R. D., and R. C. Mulligan. 1984. High-efficiency gene transfer into mammalian cells: generation of helper-free recombinant retrovirus with broad mammalian host range. Proc. Natl. Acad. Sci. USA 81:6349–6353.

11. Donoghue, D. J. 1982. Demonstration of biological activity and nucleotide sequence of an in vitro synthesized clone of the Moloney murine sarcoma virus *mos* gene. J. Virol. 42:538–546.

12. Donoghue, D. J., and T. Hunter. 1982. Expression of the transforming region of Moloney murine sarcoma virus in *E. coli* as a fusion protein with the small-t antigen of polyoma virus. Proc. Natl. Acad. Sci. USA 79:800–804.

13. Donoghue, D. J., and T. Hunter. 1983. Recombinational junctions of variants of Moloney murine sarcoma virus: generation and divergence of a mammalian transforming gene. J. Virol.

45:607–617.

14. Donoghue, D. J., P. A. Sharp, and R. A. Weinberg. 1979. Comparative study of different isolates of murine sarcoma virus. J. Virol. 32:1015–1027.

15. Frankel, A. E., and P. J. Fischinger. 1976. Nucleotide sequences in mouse DNA and RNA specific for Moloney sarcoma virus. Proc. Natl. Acad. Sci. USA 73:3705–3709.

16. Gattoni-Celli, S., W.-L. W. Hsiao, and I. B. Weinstein. 1983. Rearranged c-*mos* locus in a MOPC 21 murine myeloma cell line and its persistence in hybridomas. Nature (London) 306:795–796.

17. Gilboa, E., M. Kolbe, K. Noonan, and R. Kucherlapati. 1982. Construction of a mammalian transducing vector from the genome of Moloney murine leukemia virus. J. Virol. 44:845–851.

18. Gilboa, E., J. Park, M. Kolbe, S. Hwang, R. Kucherlapati, K. Noonan, and H. Freeman. 1982. Transduction and expression of nonselectable genes using retrovirus-derived vectors, p. 145–151. *In* Y. Gluzman (ed.), Eukaryotic viral vectors. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

19. Graham, F. L., and A. J. Van der Eb. 1973. A new technique for the assay of infectivity of human adenovirus 5 DNA. Virology 52:456–467.

20. Grunstein, M., and D. S. Hogness. 1975. Colony hybridization: a method for the isolation of cloned DNAs that contain a specific gene. Proc. Natl. Acad. Sci. USA 72:3961–3965.

21. Hannink, M., and D. J. Donoghue. 1984. Requirement for a signal sequence in biological expression of the v-*sis* oncogene. Science 226:1197–1199.

22. Hoffmann, J. W., D. Steffen, J. Gusella, C. Tabin, S. Bird, D. Cowing, and R. A. Weinberg. 1982. DNA methylation affecting the expression of murine leukemia proviruses. J. Virol. 44:144–157.

23. Hu, S., N. Davidson, and I. M. Verma. 1977. A heteroduplex study of the sequence relationships between the RNAs of M-MSV and M-MLV. Cell 10:469–477.

24. Hwang, L.-H. S., and E. Gilboa. 1984. Expression of genes introduced into cells by retroviral infection is more efficient than that of genes introduced into cells by DNA transfection. J. Virol. 50:417–424.

25. Hwang, L.-H. S., J. Park, and E. Gilboa. 1984. Role of intron-contained sequences in formation of Moloney murine leukemia virus *env* mRNA. Mol. Cell. Biol. 4:2289–2297.

26. Joyner, A. L., and A. Bernstein. 1983. Retrovirus transduction: generation of infectious retroviruses expressing dominant and selectable genes is associated with in vivo recombination and deletion events. Mol. Cell. Biol. 3:2180–2190.

27. Kloetzer, W. S., S. A. Maxwell, and R. B. Arlinghaus. 1983. p85*gag-mos* encoded by ts110 Moloney murine sarcoma virus has an associated protein kinase activity. Proc. Natl. Acad. Sci. USA 80:412–416.

28. Legerski, R. J., J. L. Hodnett, and H. B. Gray. 1978. Extracellular nucleases of Pseudomonas Bal 31. III. Use of the double-strand deoxyriboexonuclease activity as the basis of a convenient method for the mapping of fragments of DNA produced by cleavage with restriction enzymes. Nucleic Acids Res. 5:1445–1464.

29. Mann, R., R. C. Mulligan, and D. Baltimore. 1983. Construction of a retrovirus packaging mutant and its use to produce helper-free defective retrovirus. Cell 33:153–159.

30. Maxam, A. M., and W. Gilbert. 1980. Sequencing end-labeled DNA with base-specific chemical cleavages. Methods Enzymol. 65:499–560.

31. Miller, A. D., D. J. Jolly, T. Friedmann, and I. M. Verma. 1983. A transmissible retrovirus expressing human hypoxanthine phosphoribosyltransferase (HPRT): gene transfer into cells obtained from humans deficient in HPRT. Proc. Natl. Acad. Sci. USA 80:4709–4713.

32. Moloney, J. B. 1966. A virus-induced rhabdomyosarcoma of mice. Natl. Cancer Inst. Monogr. 22:139–142.

33. Mulligan, R. C. 1982. Development of new mammalian transducing vectors, p. 133–137. *In* Y. Gluzman (ed.), Eukaryotic viral vectors. Cold Spring Harbor Laboratory, Cold Spring

Harbor, N.Y.
34. Papkoff, J., T. Hunter, and K. Beemon. 1980. In vitro translation of virion RNA from Moloney murine sarcoma virus. Virology 101:91–103.
35. Papkoff, J., M. H.-T. Lai, T. Hunter, and I. M. Verma. 1981. Analysis of transforming gene products from Moloney murine sarcoma virus. Cell 27:109–119.
36. Papkoff, J., E. A. Nigg, and T. Hunter. 1983. The transforming protein of Moloney murine sarcoma virus is a soluble cytoplasmic protein. Cell 33:161–172.
37. Papkoff, J., I. M. Verma, and T. Hunter. 1982. Detection of a transforming gene product in Moloney murine sarcoma virus transformed cells. Cell 29:417–426.
38. Perkins, A. S., P. T. Kirschmeier, S. Gattoni-Celli, and I. B. Weinstein. 1983. Design of a retrovirus-derived vector for expression and transduction of exogenous genes in mammalian cells. Mol. Cell. Biol. 3:1123–1132.
39. Shinnick, T. M., R. A. Lerner, and J. G. Sutcliffe. 1981. Nucleotide sequence of Moloney murine leukaemia virus. Nature (London) 293:543–548.
40. Shtivelman, E., R. Zakut, and E. Canaani. 1984. Frequent generation of nonrescuable reorganized Moloney murine sarcoma viral genomes. Proc. Natl. Acad. Sci. USA 81:294–298.
41. Sorge, J., D. Wright, V. D. Erdman, and A. E. Cutting. 1984. Amphotropic retrovirus vector system for human cell gene transfer. Mol. Cell. Biol. 4:1730–1737.
42. Stacey, A., C. Arbuthnott, R. Kollek, L. Coggins, and W. Ostertag. 1984. Comparison of myeloproliferative sarcoma virus

with Moloney murine sarcoma virus variants by nucleotide sequencing and heteroduplex analysis. J. Virol. 50:725–732.
43. Tabin, C. J., J. W. Hoffmann, S. P. Goff, and R. A. Weinberg. 1982. Adaptation of a retrovirus as a eucaryotic vector transmitting the herpes simplex virus thymidine kinase gene. Mol. Cell. Biol. 2:426–436.
44. Temin, H. M. 1982. Function of the retrovirus long terminal repeat. Cell 28:3–5.
45. Van Beveren, C., J. A. Galleshaw, V. Jonas, A. J. M. Berns, R. F. Doolittle, D. J. Donoghue, and I. M. Verma. 1981. Nucleotide sequence and formation of the transforming gene of a mouse sarcoma virus. Nature (London) 289:258–262.
46. Van Beveren, C., F. van Straaten, J. A. Galleshaw, and I. M. Verma. 1981. Nucleotide sequence of the genome of a murine sarcoma virus. Cell 27:97–108.
47. Vande Woude, G. F., M. Oskarsson, W. L. McClements, L. W. Enquist, D. G. Blair, P. J. Fischinger, J. V. Maizel, and M. Sullivan. 1980. Characterization of integrated Moloney sarcoma provirus and flanking sequences cloned in bacteriophage lambda. Cold Spring Harbor Symp. Quant. Biol. 44:735–746.
48. Williams, D. A., I. R. Lemischka, D. G. Nathan, and R. C. Mulligan. 1984. Introduction of new genetic material into pluripotent haemotopoietic stem cells of the mouse. Nature (London) 310:476–480.
49. Wood, T. G., M. L. McGeady, D. G. Blair, and G. F. Vande Woude. 1983. Long terminal repeat enhancement of v-mos transforming activity: identification of essential regions. J. Virol. 46:726–736.

Exhibit 10

Features of the pp60^{v-src} Carboxyl Terminus That Are Required for Transformation

PETER YACIUK AND DAVID SHALLOWAY*

Department of Molecular and Cell Biology, The Pennsylvania State University, University Park, Pennsylvania 16802

Analysis of the biological and biochemical activities of pp60$^{recombinant-src}$ proteins encoded by 12 carboxyl-terminal mutants showed that a wide family of alternate *src* carboxyl termini permit complete transforming and kinase activities. *src* proteins having carboxyl termini which are up to 10 amino acids longer than that of pp60^{c-src} (17 amino acids longer than that of pp60^{v-src}) still permit transformation. Transformation-positive mutations preserve leucine-516, a residue which is highly conserved in protein-tyrosine kinase sequences; removal causes in vivo protein instability. Successive deletion mutants show that this residue is at the boundary of a region required for kinase activity. pp60^{src} which is truncated just outside this point still transforms cells and binds both pp50 and pp90 cellular proteins.

The v-*src* retroviral oncogene arose by transduction and modification of the c-*src* cellular gene (for a review, see reference 3). Unlike the viral product pp60^{v-src}, the cellular product pp60^{c-src} does not induce transformation even when overexpressed in fibroblasts (23, 40, 50) although it can induce focus formation at high expression levels (24). This may be due to the reduced protein-tyrosine kinase activity of pp60^{c-src} relative to pp60^{v-src} (7, 10, 22). The catalytic domain for this activity resides in the carboxyl half of the molecule which is strongly homologous to domains in the other sequenced protein-tyrosine kinases (for a review, see reference 21). Two cellular phosphoproteins, pp90 (a major cytoplasmic heat shock protein [38]) and pp50, specifically interact with pp60^{v-src} (4) much more extensively than with pp60^{c-src} (22).

Structurally, the viral and cellular proteins differ by the substitution of 12 different carboxyl-terminal amino acids in pp60^{v-src} for the 19 carboxyl-terminal amino acids of pp60^{c-src} and, depending on the strain of v-*src*, 8 to 15 isolated substitutions (11, 46, 59) scattered throughout the remainder of the proteins. In pp60^{c-src}, this carboxyl-terminal region contains the major site of in vivo tyrosine phosphorylation (6) which may play a role in negatively regulating pp60^{c-src} protein-tyrosine kinase activity (8). Most of the DNA sequence encoding the v-*src* carboxyl terminus is found in the chicken genome about 900 base pairs (bp) downstream from the c-*src* termination codon (57, 59).

The functional significance of the existence of multiple mutations between v-*src* and c-*src* has not yet been resolved. Single point mutations in pp60^{c-src} enable it to transform chicken embryo cells (J. B. Levy, H. Iba, and H. Hanafusa, Proc. Natl. Acad. Sci. USA, in press), and chimeric genes encoding the amino region of pp60^{c-src} and carboxyl region of pp60^{v-src} (c/v-*src* chimeras) or encoding the amino region of pp60^{v-src} and the carboxyl region of pp60^{c-src} (v/c-*src* chimeras) both induce foci in chicken or mouse cells (23, 50, 65). However, the transforming activities of these v/c-*src* genes are restricted in NIH 3T3 cells in which, in contrast with c/v-*src* chimeras, plasmid-mediated gene expression confers only weak anchorage-independent growth and tumorigenicity in syngeneic animals to the recipient cells (E. P. Reddy et al., manuscript in preparation). In addition, v/c-*src* chimeras

containing Schmidt-Ruppin D (SR-D) strain v-*src* and the c-*src* carboxyl-terminal regions have highly reduced focus-forming activities relative to unmodified SR-D v-*src* genes when tested in NIH 3T3 cells (49).

We studied the carboxyl-terminal region using a series of deletion, substitution, and addition mutants of SR-D pp60^{v-src}. We report that a wide variety of SR-D pp60^{v-src} carboxyl-terminal mutants can transform fibroblasts efficiently and completely. In addition, we show that the abrogation of transforming activity generated by previously studied pp60^{v-src} carboxyl-terminal mutants (41, 65) is due to replacement of leucine-516 and that removal of this residue, which is highly conserved among the protein-tyrosine kinases, results in pp60^{src} instability in vivo.

MATERIALS AND METHODS

Plasmid constructions. Plasmids were constructed by standard recombinant DNA techniques (31). p*src*11 was a gift from G. Cooper and A. Zelenetz.

pRS4 was derived from p*src*11 by a three-fragment ligation which resulted in the replacement of the 1,765-bp p*src*11 *Pvu*II-*Eco*RI sequence with a 1,724-bp *Pvu*II-*Eco*RI sequence originally derived from simian virus 40 (SV40) (the DNA was physically purified from pSV3*gpt* [34]). The inserted SV40 sequence encodes the 11 terminal amino acids of pRS4 (in a reading frame not used in SV40) and contains the SV40 early-region polyadenylation site (60). pRS4 has a unique *Hind*III site immediately upstream of the pRS4-*src* termination codon.

Plasmids pRS4-9, pRS4-11, pRS4-27, pRS4-18, pRS4-19, pRS4-14, and pRS4-28 were isolated from a shotgun ligation of pSV40RI *Hind*III fragments into the unique *Hind*III site of pRS4. pSV40RI, generously provided by R. Frisque (13), is a clone of *Eco*RI-linearized SV40 into the pBR322 *Eco*RI site with the SV40 late region and pBR322 ampicillin resistance gene in the same orientation. The inserted fragments (identified in Fig. 1c) are equivalent to SV40 *Hind*III fragments except that fragment A (using conventional SV40 fragment notation [60]) contains 31 bp from pBR322 between its *Eco*RI and *Hind*III sites instead of the corresponding 69 bp from the SV40 *Hind*III A fragment. Fragments identified by letters without primes are oriented such that their SV40 coordinates increase from left to right. Fragments with primes (') are inserted in the opposite orientation.

---

* Corresponding author.
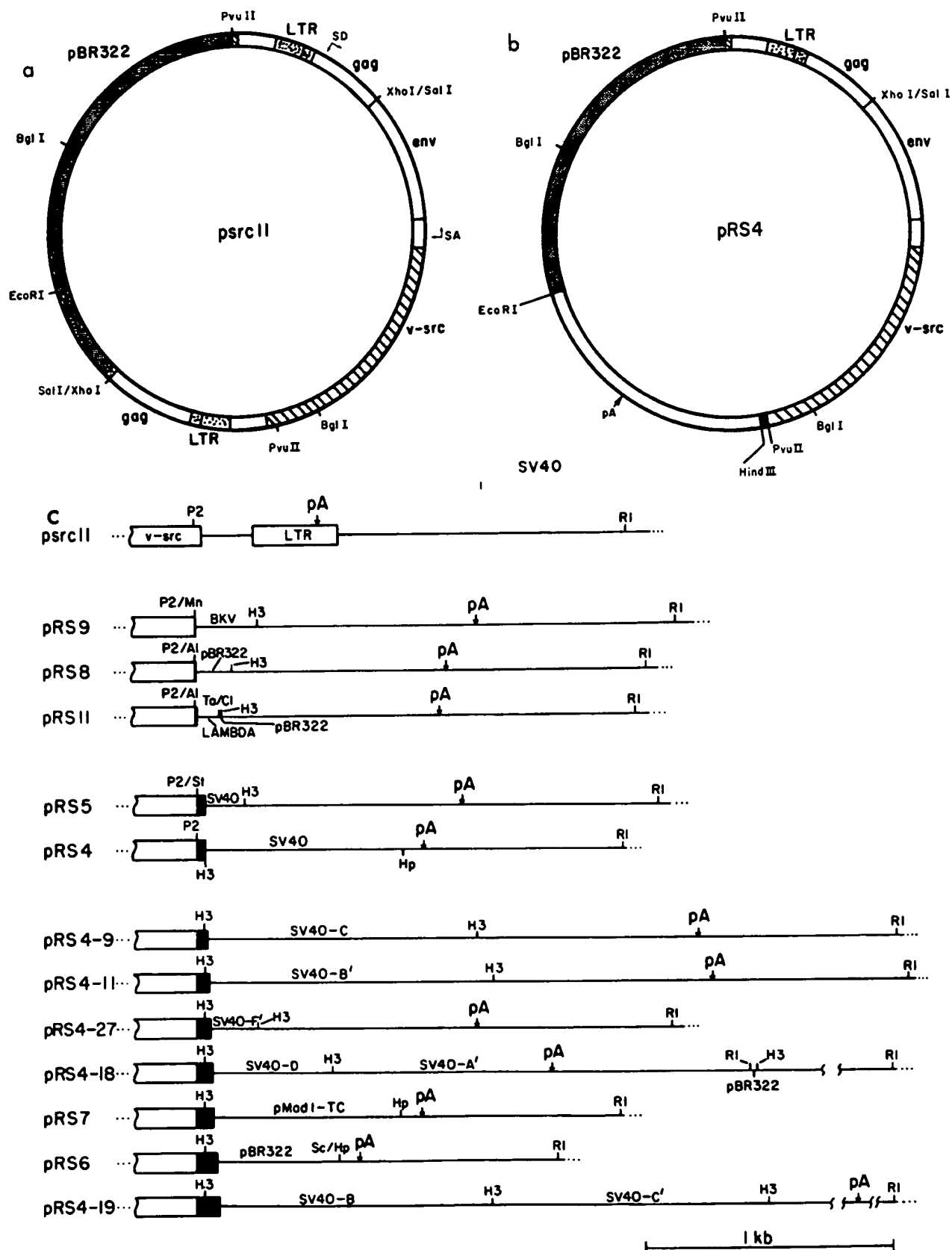
pRS4-28 (not shown) has the SV40 HindIII D, D, and C' fragments inserted in that order into the pRS4 HindIII site. It encodes the same protein as pRS4-18, but its polyadenylation site is 1,644 bp further downstream than the one in pRS4-18. pRS4-28 induced foci with the same efficiency as pRS4-18 (unpublished data), suggesting that it was unlikely that the smaller variations in polyadenylation site location in the other constructs had affected their focus-forming efficiencies.

pRS7 was constructed by ligating the 804-bp HindIII-HpaI pBR322-JC virus chimeric fragment gel purified from pMad1-TC (the fragment contains pBR322 nucleotides 31 to 0 [42, 55] and JC virus nucleotides 1724 to 951 [14]) between the HindIII and HpaI sites of pRS4-14, a plasmid identical to pRS4 (Fig. 1c) in the included region of DNA. pRS6 was constructed by a parallel insertion of the 546-bp pBR322 HindIII-ScaI fragment (pBR322 nucleotide positions 31 to 3846 [42, 55]) at the same location.

The constructions of pRS5, pRS9, pRS8, and pRS11 resulted in replacement of the 30-bp PvuII-HindIII fragment of pRS4 (Fig. 1c) with gel-purified fragments from other sources. pRS4-14 (see above) was used to gel purify the pRS4-homologous fragment and contributed an additional 6-bp pBR322 ClaI-HindIII fragment adjacent to the pRS4 HindIII site in the construction of pRS11. The inserted fragments were: pRS5, 188-bp SV40 StuI-HindIII fragment (nucleotide postions 1234 to 1046 [60]); pRS9, 262-bp BK virus MnlI-HindIII fragment (nucleotide positions 4385 to 4638 [48]; this fragment was purified from pBKV-9), a generous gift of R. Frisque [Frisque et al., submitted for publication]); pRS8, 131-bp pBR322 AluI-HindIII fragment (nucleotide positions 31 to 162 (42, 55); pRS11, 102-bp lambda cI857 AluI-TaqI fragment (nucleotide positions 31055 to 31157 [16]). Recombinants were screened and identified by multiple restriction enzyme digests.

Tissue culture and tissue culture assays. Transfection and growth of NIH 3T3 mouse cells (50) and procedures for isolating focus-selected and Eco-gpt-coselected cell lines (24) have been previously described. Coselection with pSV2neo was as described for Eco-gpt coselection, except the normal medium was supplemented with 400 μg of G-418 (Geneticin; GIBCO Laboratories, Grand Island, N.Y.) per ml for isolation of colonies and 200 μg/ml for maintenance of established cell lines (53). Polyclonal mass cell cultures were selected by growing cells from 100-mm tissue culture plates containing 50 to 70 colonies to confluence in the presence of 400 μg of G-418 per ml.

Immunoprecipitations. Cells were metabolically labeled (see below) and lysed in 0.5 ml of RIPA buffer (1% Triton X-100, 1% sodium deoxycholate, 0.1% sodium dodecyl sulfate [SDS], 150 mM NaCl, 20 mM NaH$_2$PO$_4$) supplemented with 1 mM phenylmethylsulfonyl fluoride, 2 mM EDTA, 50 mM NaF, 0.2 mM Na$_3$VO$_4$, and 100 KIU of aprotinin (Sigma Chemical Co., St. Louis, Mo.) per ml. This RIPA buffer is a good phosphatase inhibitor (15). EDTA,

NaF, and Na$_3$VO$_4$ were included to further inhibit protein kinases and phosphatases. Lysates were clarified at 25,000 × g for 30 min. Portions containing equal amounts of trichloroacetic acid (TCA)-insoluble radioactivity were adjusted to constant volume and preabsorbed twice with 50 μl of a 10% suspension of fixed protein A-containing Staphylococcus aureus bacteria for 5 min at 0°C. Proteins were immunoprecipitated in an excess of antibody from preabsorbed lysates with 1 μl of monoclonal antibody 327 (29) for 45 min at 0°C. Immune complexes were collected on 30 μl of 10% S. aureus suspension that had been precoated with 1 μg of anti-mouse immunoglobulin G (heavy plus light chains) (Miles Laboratory. Inc., Naperville, Ill.) by 20 min of incubation at 0°C, washed once with high-salt buffer (1 M NaCl, 0.5% Triton X-100, 10 mM Tris hydrochloride [pH 7.2]), and washed twice with RIPA buffer. The washed pellets were suspended in 25 μl of electrophoresis sample buffer and analyzed by electrophoresis on 7.5% SDS-polyacrylamide gels with the Laemmli (28) buffer system. [^{35}S]methionine-labeled proteins were detected by fluorography of dried gels after treatment with En^{3}Hance (New England Nuclear Corp., Boston, Mass.). ^{32}P-labeled proteins were detected by autoradiography of dried gels with Du Pont Cronex Lightning-Plus intensifying screens.

Protein kinase specific activity assays. Cells were labeled for 40 h with [^{35}S]methionine as described below. Proteins were immunoprecipitated as described above except that the immune complexes bound to protein A-containing S. aureus bacteria were washed twice with 1 ml of RIPA buffer, twice with 1 ml of high-salt buffer, suspended in 1 ml of phosphorylation buffer (20 mM Tris hydrochloride [pH 7.2], 5 mM MnCl, 2 mM 2-mercaptoethanol), and split into two 450-μl portions. After low-speed centrifugation, the pellets were resuspended in 20-μl volumes of phosphorylation buffer containing 0.1 mg of rabbit muscle enolase (Sigma) per ml. One portion from each sample was supplemented with 1 μM [γ-^{32}P]ATP (specific activity diluted to 750 Ci/mmol; New England Nuclear Corp.). Both portions were incubated at 23°C for 15 min, washed, and analyzed by SDS-polyacrylamide gel electrophoreses as described by Coussens et al. (10). [^{35}S]methionine-labeled proteins in the samples which did not contain [γ-^{32}P]ATP were detected by fluorography as described above. ^{32}P-labeled proteins were detected by autoradiography as described above except that two sheets of aluminum foil were placed between the dried gel and the X-ray film to block the ^{35}S radiation. Control experiments showed that over 95% of the film exposure with this arrangement was due to radiation emitted from ^{32}P.

Metabolic labeling. Cells were plated at 10^6 cells per 35-mm plate 24 h before labeling. For short-term [^{35}S]methionine labeling, cells were incubated for 1 h in methionine-free minimal essential medium (GIBCO) plus 10% calf serum and then incubated in 0.33 ml of methionine-free medium plus 10% calf serum and 300 μCi of [^{35}S]methionine (>1,000 Ci/mmol; New England Nuclear Corp.) per ml for 1 h. Cells

---

FIG. 1. r-src and v-src expression plasmids. All modified src genes are expressed with identical upstream regions containing Rous sarcoma virus long terminal repeats (RSV LTR), splice donor (SD) and splice acceptor (SA) sequences, and downstream sequences containing the SV40 early-region polyadenylation site (pA). psrc11 has the same upstream region but uses a downstream Rous sarcoma virus long terminal repeat for polyadenylation. (a) SR-D v-src expression plasmid psrc11. (b) r-src expression plasmid pRS4. psrc11 and pRS4 are both 7.8 kilobases (kb) long. Large dots, Rous sarcoma virus long terminal repeat; small dots, pBR322 vector sequence; hatched, v-src coding sequence; solid, nine-amino acid coding sequence substitution. (c) Modified regions of r-src DNAs. Only the modified 3' regions are shown; all other plasmid regions are identical to those shown in panels a and b. The sources of the DNAs used in the recombinant constructions are indicated (also see Materials and Methods). SV40 HindIII fragments are identified by suffixes (primes denote fragments inserted in opposite orientation to the conventional SV40 map (60). Al. AluI; Bg. BglI; Cl. ClaI; H3. HindIII; Hp. HpaI; H2. HpaII; Kp. KpnI; Mn. MnlI; P2. PvuII; Ps. PstI; RI. EcoRI; Sc. ScaI.

TABLE 1. Transforming activities of pp60^{src} Carboxyl-terminal mutants

Plasmid	Mutation	Relative focus-forming activity[a]	Growth of coselected expresser cells in soft agarose[b]
psrc11	v-src	1.0	+ +
pRS9	Deletion (Leu-516 to COOH)	<0.0005	−
pRS8	Deletion (Leu-517 to COOH)	0.2	+ +
pRS11	Deletion (Pro-518 to COOH)	0.3	+ +
pRS5	Substitution (Leu-516 to COOH)	<0.0005	ND
pRS4	Substitution (Pro-518 to COOH)	1.1	+ +
pRS4-9	Addition (4 aa[c] to pp60$^{RS4-src}$)	0.9	ND
pRS4-11	Addition (4 aa to pp60$^{RS4-src}$)	0.8	ND
pRS4-27	Addition (5 aa to pp60$^{RS4-src}$)	1.3	ND
pRS4-18	Addition (12 aa to pp60$^{RS4-src}$)	1.1	ND
pRS7	Addition (14 aa to pp60$^{RS4-src}$)	1.0	+
pRS6	Addition (17 aa to pp60$^{RS4-src}$)	0.3	+
pRS4-19	Addition (18 aa to pp60$^{RS4-src}$)	0.002	+/−

[a] All values are averages from at least three experiments of focus-forming activities relative to psrc11. The average psrc11 focus-forming activity was 2.6 × 10^4 foci per pmol.

[b] Multiple cloned cell lines or mass-culture cells that had been cotransfected with each r-src plasmid and either neo or Eco-gpt plasmids were coselected, tested for pp60^{src} expression by immunoprecipitation (e.g.. see Fig. 4), and assayed for anchorage-independent colony formation in medium containing 0.3% agarose. + +, Phenotype displayed in psrc11 panel of Fig. 3 (10 to 50% colony formation; average colony size after 14 days of 0.34 mm); +, intermediate colony formation (10 to 50% colony formation; average colony size of 0.17 mm); −, phenotype displayed in NIH 3T3 panel (0% colony formation) of Fig. 3. ND, Not done.

[c] aa, Amino acids.

for long-term [^{35}S]methionine labeling were incubated in 1.0 ml of methionine-free minimal essential medium plus 5% complete Dulbecco modified Eagle medium (GIBCO), 10% calf serum, and 100 μCi of [^{35}S]methionine per ml for 14 to 40 h. Cells were labeled with ^{32}P by incubation in 80% phosphate-free minimal essential medium–20% normal Dulbecco modified Eagle medium plus 5% calf serum for 8 h followed by incubation in phosphate-free medium containing 5% serum and 0.5 mCi of ^{32}P (ICN Radiochemicals, Inc., Irvine, Calif.) per ml for 12 to 14 h.

Determination of relative protein synthesis rates, equilibrium levels, and stabilities. Relative pp60^{src} synthesis rates were determined from the ratios of radioactivities present in pp60^{src} bands immunoprecipitated from cell lysates containing equal amounts of TCA-precipitable radioactivity after a short (1 h) [^{35}S]methionine labeling pulse (see above and Fig. 5). This period is long enough for the intracellular amino acid pool to come to equilibrium (45) but short compared with the half-lives of the pp60^{v-src} proteins (previously estimated to be 8 h [22, 47, 68]). Relative pp60^{src} equilibrium levels were determined from similar comparisons by using cells labeled for a long period (27 h) compared with the pp60^{src} half-life.

Extremely long labeling periods were avoided to reduce the possibility of metabolic perturbations from the reduced methionine concentrations required for labeling. All the cell cultures were plated and grown in parallel under identical conditions for the 48-h period before lysis except that [^{35}S]methionine was added at either 27 or 1 h before parallel culture lysis and immunoprecipitation.

For steady-state conditions, the protein synthesis rate per cell (k_s), equilibrium level per cell (p_{eq}), and turnover rate constant (k_d) are related by $p_{eq} = k_s/k_d$. [For growing cells this must be modified to $p_{eq} = k_s/(g + k_d)$ where $g = \ln 2$/cell doubling time (e.g., see reference 27), but since the pp60^{src} half-life is much less than the cell doubling time (about 24 h) $g \ll k_d$ and can be ignored.] Although the absolute values of these constants were not determined, the relative turnover rates of the pp60$^{recombinant-src}$ variants relative to that of pp60^{v-src} were determined from

$$\frac{k_d^{r-src}}{k_d^{v-src}} = \frac{k_s^{r-src}}{k_s^{v-src}} \div \frac{P_{eq}^{r-src}}{P_{eq}^{v-src}}$$

Relative stabilities were defined as the inverses of relative turnover rates.

RESULTS

The specific pp60^{v-src} carboxyl terminus is not required for transformation. Plasmid psrc11 expresses SR-D pp60^{v-src} and efficiently transforms NIH 3T3 mouse cells (50; G. Cooper and A. Zelenetz, personal communication). To determine whether the specific amino acids near the carboxyl terminus of pp60^{v-src} were required for focus-forming activity, psrc11 was modified to encode a protein in which the nine terminal amino acids of pp60^{v-src} were replaced by nine unrelated amino acids. Transfection of this plasmid, pRS4 (Fig. 1 and 2), into NIH 3T3 cells induced wild-type focus formation (Table 1). Foci and cells transformed by pRS4 were similar in morphology to those transformed by psrc11 and displayed wild-type growth in soft agarose and in vivo tumorigenicity in adult and newborn NFS mice (Table 2). In addition, pRS4

TABLE 2. In vivo tumorigenicity of pp60^{src} carboxyl-terminal mutants[a]

Cell line	No. of tumors/no. of animals injected
NIH3T3	0/9
C57(psrc11)	2/4
NIH(pRS4).G	3/7
NIH(pRS4).H	6/8
NIH(pRS4).I	6/7
C57(pRS4)	2/4
NIH(pRS8).K	5/5
NIH(pRS8).L	5/5
NIH(pRS11).K	2/7,4/7
NIH(pRS11).L	4/4

[a] Tumorigenicity was tested by injecting 1 × 10^6 NIH 3T3 cells or 5 × 10^6 C57BL/6J cells subcutaneously into newborn NFS mice or adult C57BL/6J mice, respectively. All tumors appeared within 2 to 3 weeks with the exception of one C57(pRS4) tumor that appeared within 30 days. No tumors appeared in NFS newborn mice injected with normal NIH 3T3 cells or C57BL/6J adult mice injected with normal C57BL/6J cells within the observation period of 2 months for NIH 3T3 mice or 30 days for C57BL/6J mice. Two independent experiments were performed with the NIH(pRS11).K cell line.

Amino Acid 516
↓

                                                    FFA

c-src   ... Tyr Leu Gln Ala Phe Leu Glu Asp Tyr Phe Thr Ser Thr Glu Pro Gln Tyr Gln Pro Gly Glu Asn Leu

v-src

psrcll   ... Tyr Leu Gln Ala Gln Leu Leu Pro Ala Cys Val Leu Glu Val Ala Glu   +

r-src

pRS9   ... Tyr Leu Gln Ala Gln   −
pRS8   ... Tyr Leu Gln Ala Gln Leu   −
pRS11   ... Tyr Leu Gln Ala Gln Leu Leu   + +

pRS5   ... Tyr Leu Gln Ala Gln Pro Asn Gly Arg Gln Ser Arg Val Val Ile Leu   −
pRS4   ... Tyr Leu Gln Ala Gln Leu Leu Asn Tyr Val Gly Gly Leu   −

pRS4-9   ... Tyr Leu Gln Ala Gln Leu Leu Asn Tyr Val Gly Gly Lys Leu Phe Ala Lys Ala   +
pRS4-11 ... Tyr Leu Gln Ala Gln Leu Leu Asn Tyr Val Gly Gly Lys Leu Cys Lys Asp Gly   +
pRS4-27 ... Tyr Leu Gln Ala Gln Leu Leu Asn Tyr Val Gly Leu Leu Lys Pro Phe   +
pRS4-18 ... Tyr Leu Gln Ala Gln Leu Leu Asn Tyr Val Gly Ser Pro His Ile Ile Phe   +
pRS7   ... Tyr Leu Gln Ala Gln Leu Leu Asn Tyr Val Gly Ile Ile Gln Ala Lys Gln Leu Leu Met   +
pRS6   ... Tyr Leu Gln Ala Gln Leu Leu Asn Tyr Val Gly Ile Ile Asp Asp Lys Leu Ser Asn Met Arg Ile Ile Ala Thr Leu   +
pRS4-19 ... Tyr Leu Gln Ala Gln Leu Leu Asn Tyr Val Gly Lys Leu Pro Gly His Leu Phe Ser Phe Leu Cys Phe Phe Trp Ile Lys Ile Met Leu Leu   −

FIG. 2. Carboxyl-terminal recombinant amino acid sequences. Inferred carboxyl-terminus amino acid sequences are shown beginning with pp60^{v-src} amino acid 511. All the genes encode the SR-D pp60^{v-src} amino acid sequence upstream of leucine-516. The pp60^{v-src} carboxyl-region sequence is shown for comparison. FFA, Plasmid focus-forming activity (+, >3,000 foci per pmol; −, <60 foci per pmol; see Table 1).
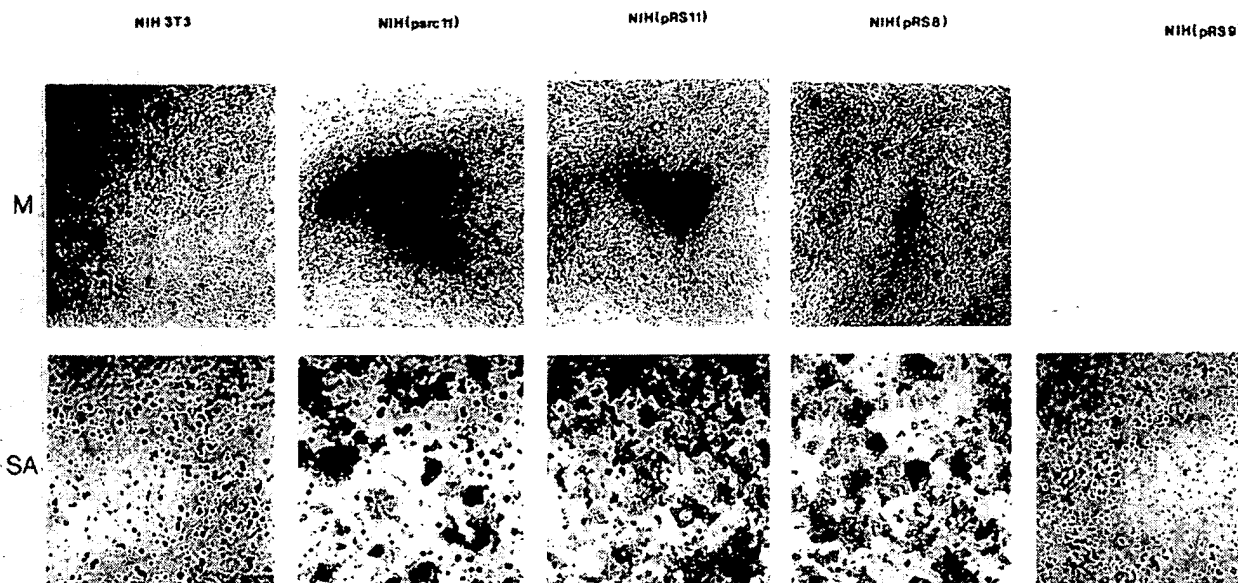
FIG. 3. Biological comparison of pp60^{src} carboxyl mutant expresser cell lines. M. Foci formed in monolayer culture after 14 days by NIH 3T3 cells transfected with plasmids psrc11, pRS11, and pRS8. No foci were observed in cells transfected with pRS9. SA. Growth in medium containing 0.3% soft agarose 12 days after suspension of G-418-resistant mass-culture cells cotransfected with pSV2neo and the indicated plasmids.
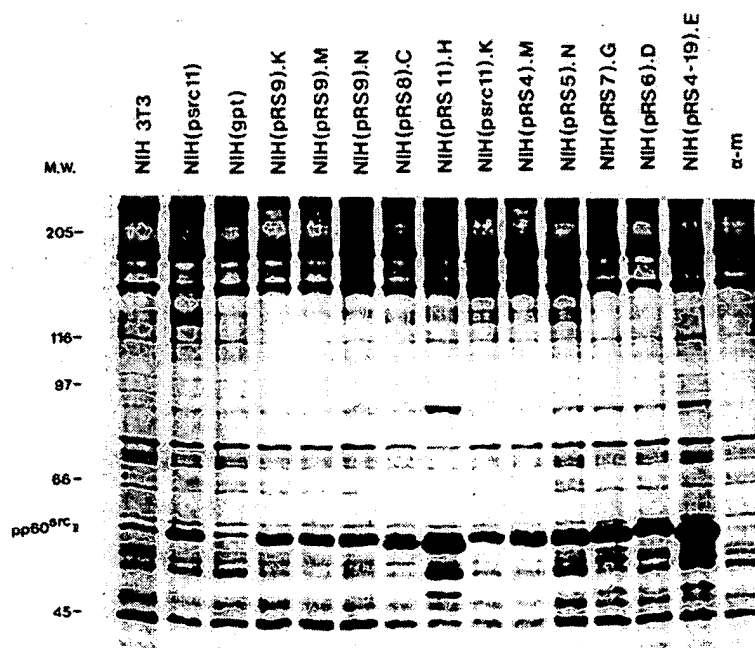
FIG. 4. Immunoprecipitation of pp60^{src} from coselected expresser cell lines. NIH 3T3 cells were cotransfected with r-src expression and coselectable marker gene plasmids. Coselected lines were labeled for 1 h with [^{35}S]methionine, and cell lysates containing equal amounts of TCA-insoluble radioactivity were immunoprecipitated with anti-src monoclonal antibody 327 (29) as described in Materials and Methods. Names within parentheses identify the transfected r-src plasmids. All cell lines were coselected with pSV302, an Eco-gpt (34) expression plasmid, except NIH(pRS9).K, which was coselected with pSV2neo (53). NIH(psrc11) is a focus-selected cell line, and NIH(gpt) is a selected cell line which was transfected only with pSV302. α-m is NIH(psrc11) lysate that was immunoprecipitated with only rabbit anti-mouse immunoglobulin G antibody and no monoclonal antibody. M.W., Molecular weight standards (×10^3). Exposure time = 4 days.
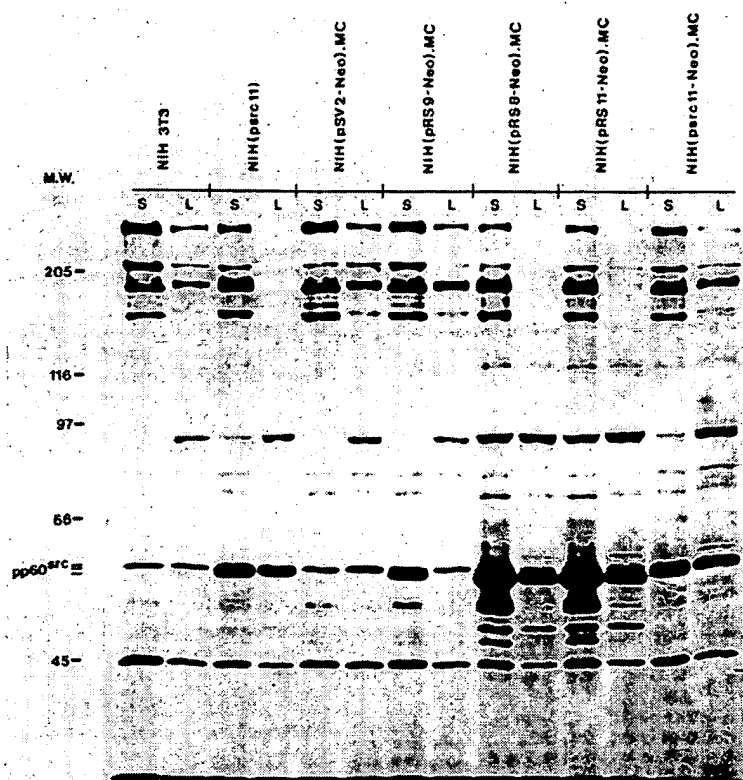
FIG. 5. Comparison of pp60$^{v\text{-}src}$ synthesis rates and equilibrium levels. Normal and mutated pp60^{src} proteins were immunoprecipitated with monoclonal antibody 327 from G-418-resistant mass-culture cells cotransfected with pSV2neo and the indicated plasmids. Immunoprecipitations were performed with cell lysates containing equal amounts of TCA-insoluble radioactivity prepared from cells labeled with [^{35}S]methionine for 1 h (S) or 27 h (L) as described in Materials and Methods. NIH(psrc11) is a focus-selected cell line not containing pSV2neo, while NIH(pSV2neo).MC is a mass-culture line transfected only with pSV2neo. M.W., Molecular weight standards (×10^3). Exposure time = 5 days.

was cotransfected with pSV302 (an Eco-gpt expression plasmid) into C57BL/6J 3T3 cells, and cells selected for Eco-gpt expression (34) were found to be tumorigenic in syngeneic mice (Table 2). Purified pRS4 DNA was tumorigenic upon direct injection into chicken wing webs (D. Robinson and H.-J. Kung, personal communication). The specific kinase activity of pp60$^{RS4\text{-}src}$ was similar to that of pp60$^{v\text{-}src}$ for both immunoglobulin G heavy chain and enolase phosphorylation (data not shown). The 3' coding sequence of pRS4 was verified by sequencing (T. Kmiecik, unpublished data).

Limits on pp60$^{v\text{-}src}$ carboxyl termini which permit transformation. Plasmids encoding 11 additional carboxyl-terminal mutants (Fig. 2) were constructed by replacing the 3' end of the v-src coding sequence with gel-purified DNA restriction fragments from numerous sources (Fig. 1) (Materials and Methods). Fragments encoding the desired 3' sequences and having appropriate restriction cleavage sites were located by computer search of sequence and plasmid data bases (51). These plasmids all contain the same upstream and downstream control sequences (Rous sarcoma virus long terminal repeat and SV40 early polyadenylation signal) as pRS4.

Three classes of recombinant-src (r-src) coding sequences were constructed: (i) elongation mutants in which additional sequence was progressively appended to the pRS4 coding sequence (pRS4-9, pRS4-11, pRS4-27, pRS4-18, pRS7, pRS6, pRS4-19), (ii) substitution mutants having the same

length as v-src (pRS4 and pRS5), and (iii) deletion mutants in which the v-src coding sequence was progressively truncated at the 3' end (pRS11, pRS8, and pRS9). Plasmid focus-forming activities were determined in multiple transfection experiments in NIH 3T3 cells (Table 1).

Progressive extension of the r-src coding sequence (Fig. 2) did not reduce focus-forming activity until the expressed protein was 17 amino acids longer than normal pp60$^{v\text{-}src}$ (pRS6) at which point a threefold reduction in efficiency was observed. Further extension (pRS4-19) almost completely eliminated focus-forming activity (Table 1).

Replacement of the last 11 (pRS5) rather than the last 9 (pRS4) amino acids of pp60$^{v\text{-}src}$ with unrelated residues eliminated focus-forming activity, suggesting that residue Leu-516 or Leu-517 or both were required for transformation. Progressive 3' truncation of the v-src coding sequence reduced focus-forming activity, but relatively high activity was observed as long as leucine-516 was preserved (pRS11 and pRS8) (Table 1). Truncation of leucine-516 and downstream residues completely eliminated focus-forming activity (pRS9). Foci induced by pRS8 and pRS11 were usually less pronounced than those induced by psrc11 (Fig. 3).

To verify that even those r-src plasmids which did not cause focus formation were inducing synthesis of the pp60$^{r\text{-}src}$ proteins, the r-src plasmids were cotransfected in a 10:1 molar ratio with plasmids expressing the neo or Eco-gpt genes, and multiple coselected cell lines were isolated with
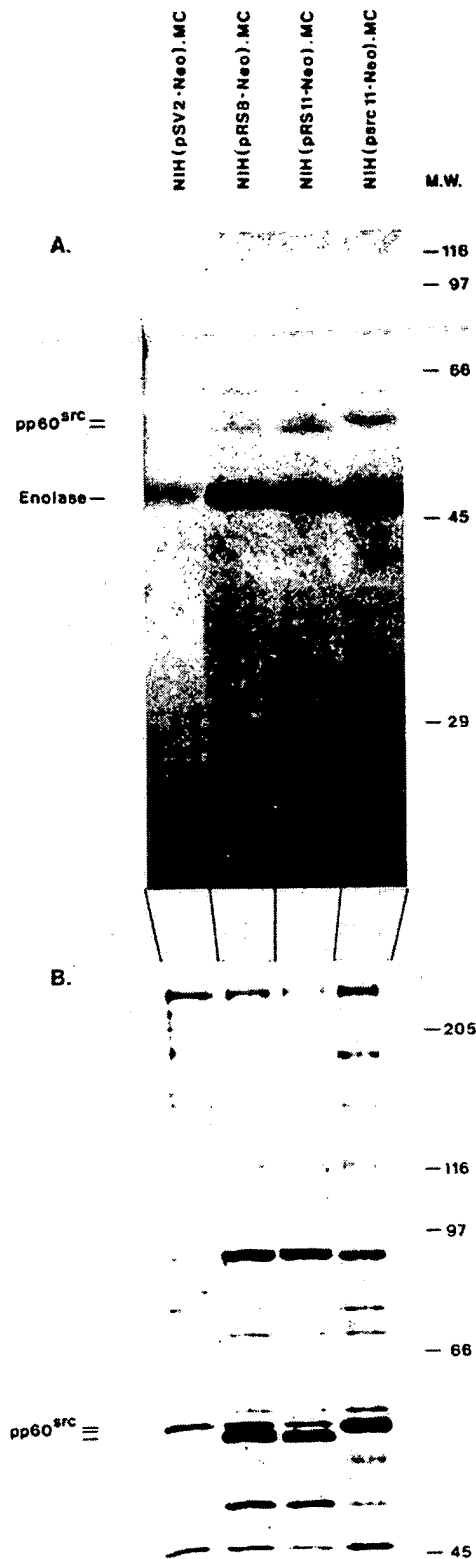
appropriate selective media (35, 53). Four or more cell lines were isolated for each *src* expression plasmid tested except for pRS5 and pRS7 for which only two cell lines each were isolated; expression plasmids pRS4-9, pRS4-11, pRS4-27, and pRS4-18 were not tested in this manner. The *src* proteins were immunoprecipitated from lysates of [^{35}S]methionine-labeled cells by using monoclonal antibody 327 (29); we expected this antibody to bind the r-*src* proteins independently of their carboxyl termini since it efficiently precipitates both pp60^{v-src} and pp60^{c-src} (30). These experiments showed that pp60^{r-src} was being efficiently synthesized in the different cell lines (Fig. 4).

The coselected r-*src* expresser cell lines were tested for colony formation in medium containing soft agarose (Table 1). Anchorage-independent growth was completely correlated with focus formation except for cells expressing the longest r-*src* protein (pp60$^{RS4-19-src}$). Although pRS4-19 did not induce foci, two of four coselected NIH(pRS4-19) lines formed colonies in soft agarose.

Requirement of leucine-516 for *src* protein stability. Mass cultures of cells expressing the truncated *src* proteins were selected with G-418 after cotransfection of pRS11, pRS8, and pRS9 with pSV2*neo*. Mass cultures selected after cotransfection with p*src*11 and pSV2*neo* and after pSV2*neo* transfection alone were used for positive and negative controls. Each culture contained cells from about 50 independent G-418-resistant colonies. The growth in soft agarose of these cultures is compared in Fig. 3. In contrast with the reduced focus-forming efficiency (Table 1) and induced focus size (Fig. 3) of pRS11 and pRS8 relative to p*src*11, they induced comparable anchorage-independent growth. No colonies were observed in multiple experiments with the pRS9-cotransfected mass culture cells.

pp60^{r-src} protein synthesis and equilibrium levels were compared (Fig. 5) by immunoprecipitation from cells labeled for periods which were short (1 h) or long (27 h) relative to the apparent pp60^{v-src} half-life (about 8 h for SR strains of v-*src* [22, 47, 68]). All the r-*src* proteins were synthesized at rates comparable to or higher than that of pp60^{v-src}, but the equilibrium level of pp60$^{RS9-src}$ was very low. Quantitative comparison showed that pp60$^{RS9-src}$ is at least five times less

FIG. 6. Specific kinase activities of r-*src* proteins. The specific activities of pp60^{src} and pp60^{r-src} proteins for autophosphorylation and enolase phosphorylation were determined with protein bound with monoclonal antibody 327 (29). Cells were labeled with 100 μCi of [^{35}S]methionine per ml for 40 h, and lysates containing 250 μg of TCA-precipitable total cell protein were immunoprecipitated as described in Materials and Methods. The *S. aureus*-bound immune complexes were suspended in phosphorylation buffer and split into two portions. Both aliquots were processed identically in the kinase assay except that [γ-^{32}P]ATP was included only in one set of reactions. NIH(pRS9*neo*).MC cells do not have a high enough equilibrium level of pp60$^{RS9-src}$ to permit specific activity quantitation. (A) Exogenous substrate kinase assay. The portions containing [γ-^{32}P]ATP in the reaction mixture were analyzed by 10% SDS-polyacrylamide gel electrophoresis and 4-day exposure of the dried gel with an enhancing screen. Two sheets of aluminum foil were placed between the gel and the film to prevent exposure from the ^{35}S radioactivity (see Materials and Methods). Some differences between [^{32}P]enolase band intensities are masked in this autoradiograph which is overexposed to display weaker autophosphorylation bands. (B) [^{35}S]methionine immunoprecipitation. The portions not containing [γ-^{32}P]ATP were analyzed by 7.5% SDS-polyacrylamide gel electrophoresis. The gels were treated with En^{3}Hance, dried, and exposed for 4 days without aluminum foil. M.W., Molecular weight ($\times10^3$).

M.W.

NIH(psrc11)
NIH(pSV2-Neo).MC
NIH(pRS9-Neo).MC
NIH(pRS8-Neo).MC
NIH(pRS11-Neo).MC
NIH(psrc11-Neo).MC
LA90 3T3

A

205—

116—

97—

—p90

66—

≡pp60^{src}

—p50

45—

29—

B

205—

116—

97—

—p90

66—

≡pp60^{src}

—p50

45—

29—

FIG. 7. Coprecipitation of pp50 and pp90 with pp60^{v-src} from ^{32}P- and ^{35}S-labeled cells. Sister cultures of G-418-resistant mass-culture lines and NIH(psrc11) were labeled with either ^{32}P or [^{35}S]methionine as described in Materials and Methods. LA90 3T3 cells (which have been previously used to study pp50-pp90-pp60^{src} binding [J. Brugge, personal communication]) were labeled only with ^{32}P under identical conditions. Immunoprecipitates from lysates containing equal amounts of TCA-precipitable ^{35}S or ^{32}P radioactivity were made with monoclonal antibody 327 and analyzed by separate 7.5% SDS-polyacrylamide gel electrophoresis as described in Materials and Methods. (A) ^{32}P-labeled cell lysates. Exposure time = 22 h. (B) [^{35}S]methionine-labeled cell lysates. Exposure time = 10 days. M.W., Molecular weight ($\times 10^3$).

stable than pp60^{v-src} (Table 3). pp60$^{RS11-src}$ and pp60$^{RS8-src}$ are only moderately less stable than pp60^{v-src}.

Different effects of pp60^{src} carboxyl-terminal mutation on kinase activity and pp50-pp90 binding. The kinase activities of the carboxyl deletion mutant pp60^{v-src} proteins bound in an immune complex with monoclonal antibody 327 were compared with enolase as a substrate and [γ-^{32}P]ATP as a phosphate donor. [^{35}S]methionine-labeled pp60^{v-src} was used so that specific activities could be accurately determined by double-label scintillation counting (see Materials and Methods). Removal of leucine-517, one amino acid beyond leucine-516, caused a decrease in kinase activity (compare activities of pp60$^{RS11-src}$ and pp60$^{RS8-src}$; Fig. 6 and Table 3). The specific kinase activity of pp60$^{RS9-src}$ could not be measured because of its low equilibrium level.

The molar phosphorylations of the deletion mutant proteins were compared by immunoprecipitating pp60^{v-src} proteins from sister cultures which had been labeled with either ^{32}P or [^{35}S]methionine. Comparison of the immunoprecipitated bands showed that pp60$^{RS11-src}$ and pp60$^{RS8-src}$ were phosphorylated at least as much as psrc11 (Fig. 7). The extent of pp60$^{RS9-src}$ phosphorylation could not be quantitatively determined because of its low equilibrium level.

All three deletion mutants bound pp50 and pp90 efficiently. These proteins comigrated with pp50 and pp90 bound to pp60^{v-src} in LA90 3T3 cells (generously provided by J. Brugge). The identities of the pp90 bands were confirmed by comparing S. aureus V-8 protease digests with those of pp90 from the LA90 3T3 cells. Both ^{35}S and multiple ^{32}P comparisons indicated that the truncated proteins bind at least as much and probably more of these cellular proteins on a molar basis than pp60^{v-src} (Table 3).

DISCUSSION

The sequence encoding the pp60^{v-src} carboxyl terminus apparently evolved by recombination between the c-src coding sequence and a cellular sequence found about 900 bp downstream from the c-src termination codon (57, 59). Homology of part of this cellular sequence to c-src suggests that it may have originated by DNA duplication, but it is not known whether it has ever been part of a protein-coding region. Whether or not this specific sequence plays a role in pp60^{v-src} induced transformation has been unknown. By constructing and characterizing genes encoding 12 different pp60^{v-src} carboxyl-terminal mutants, we showed that transforming activity is relatively insensitive to extensive replacement or deletion of almost the entire region involved in the v/c-src divergence (Fig. 2; Table 1).

All the modified genes were expressed from homologous plasmids having identical enhancer-promoter, amino coding, and polyadenylation regions. The use of naturally occurring DNA sequences identified by computer screening rather

TABLE 3. Comparison of r-src proteins in coselected cell cultures

Coselected mass culture cells	Relative pp60^r-src synthesis rate^a	Relative pp60^v-src equilibrium level^a	Relative pp60^r-src stability^b	Relative autophosphory-lation sp act^c	Relative enolase phosphoryla-tion sp act^c	p50/p60^d	p90/p60^d	pp50/pp60^e	pp90/pp60^e
NIH(pSV2neo).MC	NM^f	0.4	NM^f	0.5	NM^f	NM^f	NM	NM	NM
NIH(psrc11neo).MC	1	0.2	1	1	1	0.1	0.4	0.04	0.07
NIH(pRS11neo).MC	6	0.7	0.6	0.8	0.8	0.2	0.4	0.1	0.2
NIH(pRS8neo).MC	6	0.6	0.5	0.5	0.2	0.2	0.5	0.1	0.2
NIH(pRS9neo).MC	0.7	<0.2	<0.2	NM^f	NM	0.2^g	0.3^g	NM	0.3^g

a Values are from the experiment shown in Fig. 5. Relative synthesis rates and equilibrium levels were determined by comparing the amounts of radioactive pp60^r-src immunoprecipitated from cells labeled for times short (1 h) and long (27 h) relative to protein half-lives. All immunoprecipitations were performed with lysates containing equal amounts of TCA-precipitable radioactivity. All values are relative to the amount of radioactivity in the NIH(psrc11).MC pp60^v-src band.

b Relative stability (relative equilibrium level/relative synthesis rate) was determined as described in Materials and Methods. Stabilities are given relative to that of pp60^v-src.

c Values are from two experiments of the type shown in Fig. 6. Specific activities are normalized to pp60^v-src activity.

d Values are from the experiment shown in Fig. 7B and give the amounts of radioactivity found in the p50 or p90 bands relative to the amounts of radioactivity in the p60^r-src bands obtained by immunoprecipitation from [³⁵S]methionine-labeled cells (14-h label). All immunoprecipitations were performed with lysates containing equal amounts of TCA-precipitable radioactivity.

e Values are from the experiment shown in Fig. 7A which was executed and analyzed like that shown in Fig. 7B except that cells were labeled with ³²P (14 h).

f NM, Not accurately measurable.

g Because of the low intensity of the signal, the value is only approximate.

than synthesized oligonucleotides permitted significant experimental cost reduction. As expected, we found that the different pp60^r-src proteins were synthesized at roughly comparable rates in cotransfected cells (Fig. 4). All the modified proteins migrated as expected from their altered sizes except for pp60^RS19-src which migrated more rapidly than expected. This may be due to increased SDS binding to the rather hydrophobic tail of this protein.

The most obvious difference between pp60^c-src and pp60^v-src is in their sizes. We showed that the extended length of pp60^c-src does not by itself suppress transformation since genes encoding proteins which are up to 10 amino acids longer than pp60^c-src (pp60^RS6-src; Fig. 2) still induce efficient focus formation (Table 1). The focus-forming activities of the longer v-src proteins decreased monotonically with increasing length (Table 1), but it is possible that these decreases were sequence dependent. The appended peptides in these constructs may be loosely attached to the surface of the basic pp60^r-src structure and probably have higher internal flexibilities than normal protein structural regions. They may reduce transforming activity by steric hindrance of active sites.

Experiments done with both substitution (pRS4 and pRS5) and deletion (pRS11, pRS8, and pRS9) mutations of various extents (Fig. 1 and 2) showed that removal or (at least some) replacement of leucine-516 abrogates transforming activity (Table 1; Fig. 3). This is consistent with the result of Wilkerson et al. (65) who showed that mutation of the carboxyl-terminal 11 residues of pp60^v-src (a region including leucine-516) eliminates transforming activity.

As long as Leu-516 is preserved, reducing the length of the carboxyl terminus does not reduce pp50 and pp90 binding; rather, the data suggest that binding of these proteins is increased (Table 3). Since binding is transient (5), the apparent increases in specific binding might result from the observed increases in pp60^r-src turnover rates (Table 3) if the binding period was of relatively fixed duration. In this case the fixed binding period would occupy an increased fraction of the reduced protein half-life. This hypothesis would also account for the low pp50-pp90 specific binding to pp60^c-src which has a longer half-life than pp60^v-src (22). Alternatively, removing the carboxyl terminus may directly increase affinity. In either case, these results show that the carboxyl-

terminal nine amino acids of pp60^v-src are not required for pp50 and pp90 binding.

The in vitro exogenous substrate-specific kinase activity of pp60^RS11-src (which is missing the last nine amino acids of pp60^v-src) is similar to that of pp60^v-src (Fig. 6; Table 3), showing that most of the pp60^v-src carboxyl-terminal region which diverges from pp60^c-src is not required for kinase activity. However, specific kinase activity decreases abruptly with removal of one more residue (Leu-517), suggesting that this is near the C-terminal boundary of a part of the kinase catalytic domain. This is consistent with the finding of Gentry et al. (17) that antipeptide serum specific for pp60^v-src residues 498 to 512 inhibits phosphorylation of exogenous substrates, while antiserum specific for the terminal six residues does not inhibit kinase activity. It is interesting that this is at the boundary of a region of homology shared by the sequenced protein-tyrosine kinases (21). Since pp50-pp90-pp60^v-src complexes have reduced immunoglobulin G kinase activity (4), it is possible that the reduction in pp60^r-src kinase specific activity is caused by increased association with these proteins.

Relative pp60^r-src turnover rates were calculated from the relative amounts of radioactive label incorporated into the r-src proteins during periods which were short or long relative to the protein half-lives and showed that deletion of leucine-516 in pp60^RS9-src) results in a protein which is degraded at least five times faster than SR-D pp60^v-src (Table 3). Pulse-chase experiments with pp60^r-src proteins from coselected cell lines confirmed these differences in relative stabilities (data not shown), but effects due to reutilization of radioactive amino acids precluded the determination of absolute half-lives. (Studies with cultured cells have shown that over 80% of the amino acids labeled during the pulse period are reincorporated into protein synthesized during the chase period [45]; thus, half-lives measured in this type of experiment can be artificially extended [27, 43].)

The removal of leucine-516 may affect protein stability by altering protein conformation, transport, localization, or binding to other proteins. Our data do not indicate whether strict conservation of this residue is required for stability or whether replacement with other hydrophobic residues might preserve protein stability and activity. For example, pp60^RS9-src instability might be caused by the introduction of

TABLE 4. Conservation of leucine-516[a]

Protein	Amino acid sequence
Chicken c-*src*	...W R R D P E E R P T F E Y L Q A F L̲ E D Y F T S T E P Q Y Q P G E N L
Human c-*src* I	...W R R E P E E R P T F E Y L Q A F L̲ E D Y F T S T E P Q Y Q P G E N L
Human c-*src* II	...W R L E P E E R P T F E Y L Q S F L̲ E D Y F A S T E P Q Y Q P G D Q T
Xenopus c-*src*	.........................L Q A F L̲ E D Y F T A T E P Q Y Q P G D N L
Drosophila src	...W D A V P E K R P T F E F L N H Y L̲ E S F S V T S E V P Y R E V Q N
v-*src*	...W R R D P E E R P T F E Y L Q A Q L̲ L P A C V L E V A E
v-*fgr*	...W R L D P E E R P T F E Y L Q S F L̲ E D Y F N G P Q Q N
v-*yes*	...W K K D P D E R P T F E Y I Q S F L̲ E D Y F T A A E P S G Y
v-*ros*	...W A Q D P H N R P T F F Y I Q H K L̲ Q E I R H S P L C F S Y F L G D K + 15 aa
v-*fps*	...W E Y D P H R R P S F G A V H Q D L̲ I A I R K R H R
v-*fes*	...W A Y E P G Q R P S F S A I Y Q E L̲ Q S I R K R H R
v-*fms*	...W A L E P T R R P T F Q Q I C S L L̲ Q K Q A Q E D R R V P N Y T N L P + 16 aa
c-*fms*	...W A L E P T H R P T F Q Q I C S F L̲ Q E Q A Q E D R R E R D Y T N L P + 45 aa
v-*mil* (v-*mht*)	...L K K V R E E R P L F P Q I L S̲L̲ L Q H S L P K L N R S A S E P S L H + 21 aa
v-*raf*	...V K K V K E E R P L F P Q I L S̲L̲ L Q H S L P K L N R S A P E P S L H + 22aa
Drosophila ash	...W Q W D A T D R P T F K S I H H A L̲ E H M F Q V G N V
Isk^T (*tck*)	...W K E R P E D R P T F D Y L R S V L̲ D D F F T A T E G Q Y Q P Q P
Human insulin rec.	...W Q F N P N M R P T F L E I V N L L̲ K D D L H P S F P E V S
v-*mos*	...I Q S C W E A R G L Q R P S A E L L̲ Q R D L K A F R G T L G
Human EGF rec.	...W M I D A D S R P K F R E L I I E F S K M A R D P Q R Y L V I Q G D E + 225 aa
c-*erbB*-2	...W M I D S E C R P R F R E L V S E F S R M A R D P Q R F V V I Q N E D + 262 aa
v-*erbB*	...W M I D A D S R P K F R E L I A E F S K M A R D P D R Y L V I Q G D E + 199 aa
neu	...W M I D S E C R P R F R E L V S E F S R M A R D P Q R F V V I Q N E D + 263 aa
v-*abl*	...W Q W N P S D R P S F A E I H Q A F E T M F Q E S S I S D E V E K E L + 287 aa

[a] The amino acid sequence in the carboxyl-terminal region of pp60^{v-src} is aligned with the homologous sequences from 23 other protein-tyrosine kinase-related proteins encoded by chicken c-*src* (59), human c-*src* I (1, 39), human c-*src* II (39), *Xenopus* c-*src* (54), *Drosophila* src and ash (20), v-*src* (11,46,58), v-*fqr* (36), v-*yes* (26), v-*ros* (37), v-*fps* (52), v-*fes* (19), v-*fms* (18), c-*fms* (9), v-*mil* (v-*mht*) (25,56), v-*raf* (32), Isk^T (*tck*) (33,63), human insulin receptor (61), v-*mos* (62), human epidermal growth factor (EGF) receptor (12,61), c-*erbB*-2 (67), v-*erbB* (66), neu (2), and v-*abl* (44). The chicken c-*src*, v-*src*, v-*fgr*, v-*yes*, v-*ros*, v-*fps*, v-*fes*, v-*abl*, human EGF receptor, and v-*erbB* sequences were aligned as in Hunter and Cooper (21). The additional cellular *src*, c-*erbB*-2, neu, and *abl* sequences were aligned by obvious homology. Only part of the *Xenopus* c-*src* sequence has been published. The c-*fms*, v-*fms*, v-*mil*, v-*raf*, and human insulin receptor sequences were aligned by using the almost completely conserved sequence RPXF at *src* positions 506 to 509. In agreement with Mark and Rapp (32), the v-*mos* sequence was aligned at the conserved R at *src* position 506. Beyond adjustment of the amino acids SIA and SIE in the v-*mil* and v-*raf* sequences, no attempt has been made to maximize homologies by relative insertions or deletions. aa, Amino acids.

a negative charge (from the COOH terminus) into an internal region of the protein. While a complete understanding of the significance of Leu-516 may have to await determination of the pp60^{src} three-dimensional structure, comparison with the homologous domains from other sequenced protein-tyrosine kinases and related proteins (Table 4) shows that this residue is highly conserved (present in 18 of 23 sequences), particularly when compared with the conservation of its immediately neighboring residues. It is completely conserved in the proteins in which it is located near the carboxyl terminus and is invariably altered to another hydrophobic residue, phenylalanine, in the proteins in which it is located more internally in the primary sequence. Sequence comparison throughout the entire catalytic domain (21) shows that leucine-516 is the furthest downstream of the highly conserved residues.

This residue is also notable in that, in the cellular protooncogene product pp60^{c-src}, it marks the amino boundary of a region which appears to have the hallmarks of a compact loop structure (a strongly hydrophilic region preferentially containing Gly, Pro, Asp, Asn, Ser, and Tyr flanked by hydrophobic residues [J. F. Leszczynski and G. D. Rose, submitted for publication]). This potential structure extends to the carboxyl terminus of pp60^{c-src} and includes its major site of tyrosine phosphorylation (6). The observation that phosphorylation of this tyrosine lowers pp60^{c-src} kinase specific activity (8) suggests that this region of the pp60^{c-src} molecule provides some negative regulatory function. This suggestion is supported by the finding that replacement of this region in p*src*11 with the corresponding region of c-*src* in

constructs similar to those discussed here almost completely eliminates transforming activity in NIH 3T3 cells (49; P. Yaciuk, unpublished data). Thus, the primary significance of the pp60^{v-src} carboxyl terminus may be just that it eliminates the negatively acting pp60^{c-src} carboxyl terminus. From this standpoint, the fact that a wide class of modifications of the pp60^{v-src} carboxyl terminus do not eliminate transforming activity is not surprising.

LITERATURE CITED

1. Anderson, S. K., C. P. Gibbs, A. Tanaka, H.-J. Kung, and D. J. Fujita. 1985. Human cellular *src* gene: nucleotide sequence and derived amino acid sequence of the region coding for the carboxyl-terminal two-thirds of pp60^{c-src}. Mol. Cell. Biol. 5:1122–1129.
2. Bargmann, C. I., M.-C. Hung, and R. A. Weinberg. 1986. The

neu oncogene encodes an epidermal growth factor receptor-related protein. Nature (London) 319:226–230.

3. Bishop, M. J. 1983. Cellular oncogenes and retroviruses. Annu. Rev. Biochem. 52:301–354.

4. Brugge, J. S., E. Erikson, and R. L. Erikson. 1981. The specific interaction of the Rous sarcoma virus transforming protein, pp60^{src}, with two cellular proteins. Cell 25:363–372.

5. Brugge, J., W. Yonemoto, and D. Darrow. 1983. Interaction between the Rous sarcoma virus transforming protein and two cellular phosphoproteins: analysis of the turnover and distribution of this complex. Mol. Cell. Biol. 3:9–19.

6. Cooper, J. A., K. L. Gould, C. A. Cartwright, and T. Hunter. 1986. Tyr^{527} is phosphorylated in pp60^{c-src}: implications for regulation. Science 231:1431–1434.

7. Cooper, J. A., T. Hunter, and D. Shalloway. 1985. Protein-tyrosine kinase activity of pp60^{c-src} is restricted in intact cells, p. 321–328. *In* J. Feramsco, B. Ozanne, and C. Stiles (ed.), Cancer cells, vol. 3. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

8. Courtneidge, S. A. 1985. Activation of the pp60^{c-src} kinase by middle T antigen binding or by dephosphorylation. EMBO J. 4:1471–1477.

9. Coussens, L., C. V. Beveren, D. Smith, E. Chen, R. L. Mitchell, C. M. Isacke, I. M. Verma, and A. Ullrich. 1986. Structural alteration of viral homologue of receptor proto-oncogene *fms* at carboxyl terminus. Nature (London) 320:277–280.

10. Coussens, P. M., J. A. Cooper, T. Hunter, and D. Shalloway. 1985. Restriction of the in vitro and in vivo tyrosine protein kinase activities of pp60^{c-src} relative to pp60^{v-src}. Mol. Cell. Biol. 5:2753–2763.

11. Czernilosky, A. P., A. D. Levinson, H. E. Varmus, J. M. Bishop, E. Tischer, and H. Goodman. 1983. Corrections to the nucleotide sequence of the *src* gene of Rous sarcoma virus. Nature (London) 301:736–738.

12. Downward, J., Y. Yarden, E. Mayes, G. Scrace, N. Totty, P. Stockwell, A. Ullrich, J. Schlessinger, and M. D. Waterfield. 1984. Close similarity of epidermal growth factor receptor and v-*erb*-B oncogene protein sequences. Nature (London) 307:521–527.

13. Frisque, R. J. 1983. Regulatory sequences and virus-cell interactions of JC virus, p. 41–59. *In* J. L. Sever and D. L. Madden (ed.), Polyomaviruses and human neurological diseases. Alan R. Liss, Inc., New York.

14. Frisque, R. J., G. L. Bream, and M. T. Cannella. 1984. Human polyomavirus JC virus genome. J. Virol. 51:458–469.

15. Foulkes, J. G., E. Erikson, and R. L. Erikson. 1983. Separation of multiple phosphotyrosyl- and phosphoseryl-protein phosphatases from chicken brain. J. Biol. Chem. 258:431–438.

16. GenBank and European Molecular Biology Laboratory. 1985. Phage 12-24. Nucleotide sequences, vol. 3. IRL Press, Oxford, England.

17. Gentry, L. E., L. R. Rohrschneider, J. Casnellie, and E. G. Krebs. 1983. Antibodies to defined region of pp60^{src} neutralize the tyrosine specific kinase activity. J. Biol. Chem. 258:11219–11228.

18. Hampe, A., M. Gobet, C. J. Sherr, and F. Galibert. 1984. Nucleotide sequene of the feline retroviral oncogene v-*fms* shows unexpected homology with oncogenes encoding tyrosine-specific protein kinases. Proc. Natl. Acad. Sci. USA 81:85–89.

19. Hampe, A., I. Laprevotte, and F. Galibert. 1982. Nucleotide sequences of feline retroviral oncogenes (v-*fes*) provide evidence for a family of tyrosine-specific protein kinase genes. Cell 30:775–785.

20. Hoffman, F. M., L. D. Fresco, H. Hoffman-Falk, and B.-Z. Shilo. 1983. Nucleotide sequences of the *Drosophila src* and *abl* homologs: conservation and variability in the *src* family oncogenes. Cell 35:393–401.

21. Hunter, T., and J. A. Cooper. 1985. Protein-tyrosine kinases. Annu. Rev. Biochem. 54:897–930.

22. Iba, H., F. R. Cross, E. A. Garber, and H. Hanafusa. 1985. Low level of cellular protein phosphorylation by nontransforming overproduced p60^{src}. Mol. Cell. Biol. 5:1058–1066.

23. Iba, H., T. Takeya, F. R. Cross, T. Hanafusa, and H. Hanafusa.

1984. Rous sarcoma virus variants that carry the cellular *src* gene instead of the viral *src* gene cannot transform chicken embryo fibroblasts. Proc. Natl. Acad. Sci. USA 81:4424–4428.

24. Johnson, P. J., P. M. Coussens, A. V. Danko, and D. Shalloway. 1985. Overexpressed pp60^{c-src} can induce focus formation without complete transformation of NIH 3T3 cells. Mol. Cell. Biol. 5:1073–1083.

25. Kan, N. C., C. S. Flordellis, G. E. Mark, P. H. Duesberg, and T. S. Papas. 1984. Nucleotide sequence of avian carcinoma virus MH2: two potential *onc* genes, one related to avian virus MC29 and the other related to murine sarcoma virus 3611. Proc. Natl. Acad. Sci. USA 81:3000–3004.

26. Kitamura, N., A. Kitamura, K. Toyoshima, Y. Hirayama, and M. Yoshida. 1982. Avian sarcoma virus Y73 genome sequence and structural similarity of its transforming gene product to that of Rous sarcoma virus. Nature (London) 297:205–208.

27. Koch, A. L. 1962. The evaluation of the rates of biological processes from tracer kinetic data. I. The influence of labile metabolic pools. J. Theor. Biol. 3:283–303.

28. Laemmli, U. K. 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature (London) 227:680–685.

29. Lipsich, L. A., A. J. Lewis, and J. S. Brugge. 1983. Isolation of monoclonal antibodies that recognize the transforming proteins of avian sarcoma viruses. J. Virol. 48:352–360.

30. Lipsich, L. A., W. Yonemoto, J. B. Bolen, M. A. Israel, and J. S. Brugge. 1984. Structural and functional studies of the Rous sarcoma virus transforming protein, pp60^{src}, p. 43–51. *In* G. F. Vande Woude, A. J. Levine, W. C. Topp, and J. D. Watson (ed.), Cancer cells, vol. 2. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

31. Maniatis, T., E. F. Fritsch, and J. Sambrook. 1982. Molecular cloning, a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

32. Mark, G. E., and U. R. Rapp. 1984. Primary structure of v-*raf*: relatedness to the *src* family of oncogenes. Science 224:285–289.

33. Marth, J. D., R. Peet, E. G. Krebs, and R. M. Perlmutter. 1985. A lymphocyte-specific protein-tyrosine kinase gene is rearranged and overexpressed in the murine T cell lymphoma LSTRA. Cell 43:393–404.

34. Mulligan, R. C., and P. Berg. 1980. Expression of a bacterial gene in mammalian cells. Science 209:1422–1427.

35. Mulligan, R. C., and P. Berg. 1981. Selection for animal cells that express the *Escherichia coli* gene coding for xanthine-guanine phosphoribosyltransferase. Proc. Natl. Acad. Sci. USA 78:2072–2076.

36. Naharro, G., K. C. Robbins, and E. P. Reddy. 1984. Gene product of v-*fgr onc*: hybrid protein containing a portion of actin and a tyrosine-specific protein kinase. Science 223:63–66.

37. Neckameyer, W. S., and L.-H. Wang. 1985. Nucleotide sequence of avian sarcoma virus UR2 and comparison of its transforming gene with other members of the tyrosine protein kinase oncogene family. J. Virol. 3:879–884.

38. Oppermann, H., W. Levinson, and J. M. Bishop. 1981. A cellular protein that associates with a transforming protein of Rous sarcoma virus is also a heat shock protein. Proc. Natl. Acad. Sci. USA 78:1067–1071.

39. Parker, R. C., G. Mardon, R. V. Lebo, H. E. Varmus, and J. M. Bishop. 1985. Isolation of duplicated human c-*src* genes located on chromosomes 1 and 20. Mol. Cell. Biol. 5:831–838.

40. Parker, R. C., H. E. Varmus, and J. M. Bishop. 1984. Expression of v-*src* and chicken c-*src* in rat cells demonstrates qualitative differences between pp60^{v-src} and pp60^{c-src}. Cell 37:131–139.

41. Parsons, J. T., D. Bryant, V. Wilkerson, G. Gilmartin, and S. J. Parsons. 1984. Site-directed mutagenesis of Rous sarcoma virus pp60^{src}: identification of functional domains required for transformation, p. 37–42. *In* G. F. Vande Woude, A. J. Levine, W. C. Topp, and J. D. Watson (ed.), Cancer cells, vol. 2. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

42. Peden, K. W. C. 1983. Revised sequence of the tetracycline-resistance gene of pBR322. Gene 22:277–280.

43. Poole, B. 1971. The kinetics of disappearance of labeled leucine

from the free leucine pool of rat liver and its effect on the apparent turnover of catalase and other hepatic proteins. J. Biol. Chem. **246**:6587–6591.

44. **Reddy, E. P., M. J. Smith, and A. Srinivasan.** 1983. Nucleotide sequence of Abelson murine leukemia virus genome: structural similarity of its transforming gene product to other *onc* gene products with tyrosine-specific kinase activity. Proc. Natl. Acad. Sci. USA **80**:3623–3627.

45. **Righetti, P., E. P. Little, and G. Wolf.** 1971. Reutilization of amino acids in protein synthesis in HeLa cells. J. Biol. Chem. **246**:5724–5732.

46. **Schwartz, D. E., R. Tizard, and W. Gilbert.** 1983. Nucleotide sequence of Rous sarcoma virus. Cell **32**:853–869.

47. **Sefton, B. M., T. Patschinsky, C. Berdot, T. Hunter, and T. Elliott.** 1982. Phosphorylation and metabolism of the transforming protein of Rous sarcoma virus. J. Virol. **41**:813–820.

48. **Seif, I., G. Khoury, and R. Dhar.** 1979. The genome of human papovavirus BKV. Cell **18**:963–977.

49. **Shalloway, D., P. M. Coussens, and P. Yaciuk.** 1984a. c-*src* and *src* homolog overexpression in mouse cells, p. 9–17. *In* G. F. Vande Woude, A. J. Levine, W. C. Topp, and J. D. Watson (ed.), Cancer cells, vol. 2. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

50. **Shalloway, D., P. M. Coussens, and P. Yaciuk.** 1984b. Overexpression of the c-*src* protein does not induce transformation of NIH 3T3 cells. Proc. Natl. Acad. Sci. USA **81**:7071–7075.

51. **Shalloway, D., and N. R. Deering.** 1984. Recombinant DNA data management at the restriction and functional site level. Nucleic Acids Res. **12**:739–750.

52. **Shibuya, M., and H. Hanafusa.** 1982. Nucleotide sequence of Fujinami sarcoma virus: evolutionary relationship of its transforming gene with transforming genes of other sarcoma viruses. Cell **30**:787–795.

53. **Southern, P. J., and P. Berg.** 1982. Transformation of mammalian cells to antibiotic resistance with a bacterial gene under control of the SV40 early region promoter. J. Mol. Appl. Genet. **1**:327–341.

54. **Steele, R. E.** 1985. Two divergent cellular *src* genes are expressed in *Xenopus laevis*. Nucleic Acids Res. **13**:1747–1761.

55. **Sutcliffe, J. G.** 1979. Complete sequence of the *Escherichia coli* plasmid pBR322. Cold Spring Harbor Symp. Quant. Biol. **43**:77–90.

56. **Sutrave, P., T. I. Bonner, U. R. Rapp, H. W. Jansen, T. Patschinsky, and K. Bister.** 1984. Nucleotide sequence of avian retroviral oncogene v-*mil*: homologue of murine retroviral

oncogene v-*raf*. Nature (London) **309**:85–88.

57. **Swanstrom, R., R. C. Parker, H. E. Varmus, and J. M. Bishop.** 1983. Transduction of a cellular oncogene: the genesis of Rous sarcoma virus. Proc. Natl. Acad. Sci. USA **80**:2519–2523.

58. **Takeya, T., R. A. Feldman, and H. Hanafusa.** 1982. DNA sequence of the viral and cellular *src* gene of chickens. I. Complete nucleotide sequence of an *Eco*RI fragment of recovered avian sarcoma virus which codes for gp37 and pp60^{src}. J. Virol. **44**:1–11.

59. **Takeya, T., and H. Hanafusa.** 1983. Structure and sequence of the cellular gene homologous to the RSV *src* gene and the mechanism for generating the transforming virus. Cell **32**:881–890.

60. **Tooze, J.** (ed.) 1981. Molecular biology of tumor viruses, part 2: DNA tumor viruses, 2nd ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

61. **Ullrich, A., J. R. Bell, E. Y. Chen, R. Herrera, L. M. Petruzzelli, T. J. Dull, A. Gray, L. Coussens, Y.-C. Liao, M. Tsubokawa, A. Mason, P. H. Seeburg, C. Grunfeld, O. M. Rosen, and J. Ramachandran.** 1985. Human insulin receptor and its relationship to the tyrosine kinase family of oncogenes. Nature (London) **313**:756–761.

62. **Van Beveren, C., J. A. Galleshaw, V. Jonas, A. J. M. Berns, R. F. Doolittle, D. J. Donoghue, and I. M. Verma.** 1981. Nucleotide sequence and formation of the transforming gene of a mouse sarcoma virus. Nature (London) **289**:258–262.

63. **Voronova, A. F., and B. M. Sefton.** 1986. Expression of a new tyrosine protein kinase is stimulated by retrovirus promoter insertion. Nature (London) **319**:682–685.

64. **Welch, W. J., and J. R. Feramisco.** 1982. Purification of the major mammalian heat shock proteins. J. Biol. Chem. **257**:14949–14959.

65. **Wilkerson, V. W., D. L. Bryant, and J. T. Parsons.** 1985. Rous sarcoma virus variants that encode *src* proteins with an altered carboxyl terminus are defective for cellular transformations. J. Virol. **55**:314–321.

66. **Yamamoto, T., T. Nishida, N. Miyajima, S. Kawai, T. Ooi, and K. Toyoshima.** 1983. The *erbB* gene of avian erythroblastosis virus is a member of the *src* gene family. Cell **35**:71–78.

67. **Yamamoto, T., S. Ikawa, T. Akiyama, K. Semba, N. Nomura, N. Miyajima, T. Saito, and K. Toyoshima.** 1986. Similarity of protein encoded by the human c-*erb*-B-2 gene to epidermal growth factor receptor. Nature (London) **319**:230–234.

68. **Ziemiecki, A., R. R. Friis, and H. Bauer.** 1982. Half-life of the Rous sarcoma virus transforming protein pp60^{src} and its associated kinase activity. Mol. Cell. Biol. **2**:355–360.

Exhibit 11

Catalytic and non-catalytic domains of the Fujinami sarcoma virus P130gag-fps protein-tyrosine kinase distinguished by the expression of v-*fps* polypeptides in *Escherichia coli*

Ivan Sadowski & Tony Pawson

Division of Molecular and Developmental Biology, Mount Sinai Hospital Research Institute, 600 University Ave., Toronto, Ontario M5G 1X5 and Department of Medical Genetics, University of Toronto, Ontario, Canada

While protein-tyrosine kinases share a region of sequence identity corresponding to their kinase domains, the specific elements essential for catalysis, substrate binding and substrate specificity are largely undefined. The P130gag-fps transforming protein of Fujinami avian sarcoma virus is a cytoplasmic tyrosine kinase with a complex structure that includes a C-terminal kinase domain. To identify the precise N-terminal border of the v-*fps* catalytic region and to assess its interactions with non-catalytic domains, C-terminal v-*fps* polypeptide fragments of decreasing size were expressed in *E. coli* as trpE-v-*fps* hybrid proteins. All such polypeptides containing 263 or more residues derived from the C-terminus of P130gag-fps (i.e. residues 920–1182) were enzymatically active as tyrosine kinases. They autophosphorylated at physiological sites *in vivo* and phosphorylated exogenous substrates such as enolase and poly(glu,tyr) at tyrosine *in vitro*. Deletion of a further five amino acids from P130gag-fps residues 920–925 abolished all enzymatic activity. This deletion coincides with the predicted N-terminus of the v-*fps* ATP-binding site at residue 922. These data indicate that the N-terminal border of the ATP-binding site defines the start of the minimal v-*fps* tyrosine kinase catalytic domain, and show that this minimal domain is competent to bind substrates. More N-terminal non-catalytic sequences appear to functionally interact with the catalytic domain.

Introduction

Protein-tyrosine kinases (PTKs) are complex polypeptides composed of multiple structural and functional domains (Levinson *et al.*, 1981; Stoker *et al.*, 1984; Stone *et al.*, 1984; Stone & Pawson, 1985). The catalytic activities of these enzymes are apparently modified by regulatory domains with a variety of functions, including targetting of newly synthesized protein to the plasma membrane and direct modulation of kinase activity (Garber *et al.*, 1985; Pellman *et al.*, 1985; Cooper *et al.*, 1986; Sadowski *et al.*, 1986). It also seems likely that non-catalytic sequences will prove to be involved in substrate recognition (Sadowski *et al.*, 1986; Jove *et al.*, 1986). Structural alterations in one or more of these regulatory domains leading to derepressed enzymatic activity are frequently involved in the oncogenic activation of PTKs (Downward *et al.*, 1984; Shtivelman *et al.*, 1985; Yaciuk & Shalloway, 1986).

For a PTK, the precise definition of the catalytic domain and its intramolecular interactions with adjacent elements is important in understanding the mechanisms by which enzymatic activity is regulated and cellular targets are selected. We have previously employed a variety of genetic and biochemical techniques to investigate the domain structure of the P130gag-fps tyrosine kinase encoded by Fujinami avian sarcoma virus (FSV) and the functions of specific amino acids within its kinase region (Stone *et al.*, 1984; Weinmaster *et al.*, 1983, 1984, 1986). The catalytic domain of P130gag-fps can be isolated by partial proteolysis within a 29-kDa C-terminal fragment that retains enzymatic activity (Weinmaster *et al.*, 1983). In keeping with this observation, approximately 260 residues near the C-terminus of P130gag-fps (residues 911–1173) show sequence identity with all other PTKs and to a lesser extent with serine/threonine-specific protein kinases. Immediately to the N-terminus of the predicted kinase domain is a sequence of approximately 90 amino acids (residues 810–900) which is shared with other cytoplasmic tyrosine kinases such as p60src, and which may associate with cellular proteins that regulate or mediate kinase function (Sadowski *et al.*, 1986). We have termed this domain SH2 (for *src* homology 2). The proposed SH2 and catalytic domains are joined to an N-terminal *fps*-specific domain unnecessary for kinase activity but important for the induction of neoplastic transformation (Stone *et al.*, 1984; Stone & Pawson, 1985).

As a means to identifying the precise boundaries of P130gag-fps domains and to investigate their functional interactions, we have constructed plasmid vectors for the synthesis of v-*fps* polypeptides in *E. coli*. By manipulation of the coding sequences for the expressed proteins we define the N-terminal boundary for the kinase catalytic domain, and identify potential interactions with non-catalytic sequences.

Results

High level expression of v-fps polypeptides in E. coli

v-*fps* coding sequences derived from the FSV genome were expressed as trpE-v-*fps* polypeptides in *E. coli* following insertion into the multiple cloning site of pATH vectors (Figure 1). These plasmids allow the expression of foreign sequences as hybrid proteins whose N-termini are encoded by the bacterial *trpE* gene. Synthesis of fusion proteins can be induced by tryptophan starvation and growth of bacteria in the presence of indole acrylic acid. Two sets of expression plasmids were constructed as detailed in Figure 1 and in Materials and methods. Table
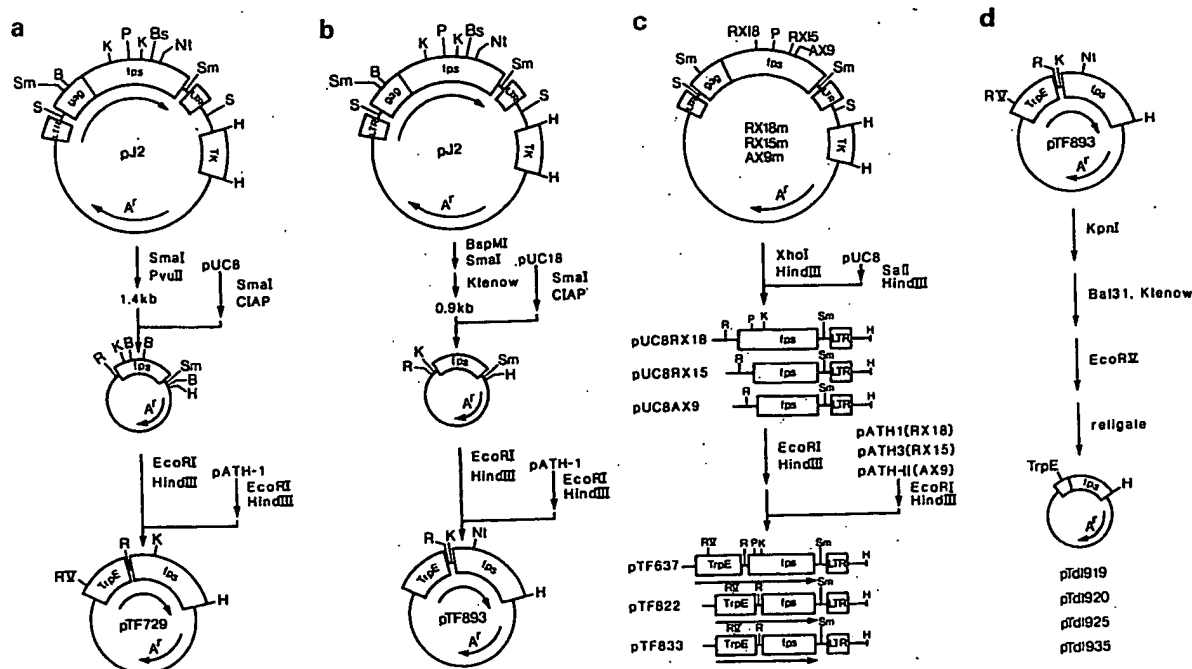
Figure 1 Construction of *trpE*-v-*fps* bacterial expression vectors. v-*fps* sequences encoding C-terminal polypeptide fragments of Fujinami avian sarcoma virus (FSV) P130^gag-fps were excised from wt pJ2 plasmid (a and b) or the RX18m, RX15m and AX9m FSV mutant plasmids containing unique XhoI sites (c). Fragments with blunt ends generated by digestion with SmaI or PvuII/SmaI were cloned into the SmaI site of pUC vectors (a). The BspMI/SmaI digested wt DNA was first made blunt by treatment with Klenow enzyme (b). C-terminal encoding fragments frame XhoI linker mutants were cloned into the SalI/HindIII sites of pUC8 (c). The various sized v-*fps* pUC inserts were then subcloned into the proper reading form of pATH-1, pATH-3 or pATH-11 with EcoRI and HindIII (a, b and c). For construction of Bal31 deletions, pTF729 (b bottom) or pTF893 (b bottom) were digested with KpnI, treated with Bal31 and Klenow enzyme and then religated to the EcoRV site of the *trpE* coding sequence (d). Abbreviations are: B = BamHI, H = HindIII, K = KpnI, P = PvuII, R = EcoRI, S = SstI, Sm = SmaI, Bs = BspMI, Nt = NotI, X = XhoI, RV = EcoRV

1 summarizes the coding potential of these *trpE*-v-*fps* vectors. One class of plasmids, denoted pTF, encode proteins with 323 N-terminal *trpE* amino acids (37 kDa) linked in-frame to a nested set of v-*fps* C-terminal sequences. A second series of plasmids designated pTdl encode fusion proteins with only 42 N-terminal *trpE* residues joined to a distinct set of v-*fps* polypeptides (see

Table 1). In each case, the residue within P130^gag-fps corresponding to the first v-*fps* amino acid in the *trpE*-v-*fps* fusion protein is indicated in the plasmid name; all fusion proteins have a comon C-terminus corresponding to the C-terminal amino acid (residue 1182) of P130^gag-fps.

Following induction, RRI *E. coli* containing the plasmid constructs were found to abundantly express new

Table 1 Coding potential of *trpE*-v-*fps* expression plasmids

Plasmid	trpE amino acids[a]	v-fps amino acids[c]	Protein molecular weight[d]	5' fps End derived from[e]
pTF49	1–323[b]	49–1182	167.3	SmaI pJ2
pTdl359	1–42	359–1182	99.4	NcoI pTF49
pTF637	1–323[b]	637–1182	99.6	XhoI pRX18m
pTF729	1–323[b]	729–1182	89.1	PvuII pJ2
pTF822	1–323[b]	822–1182	78.4	XhoI pRX15m
pTdl823	1–42	823–1182	46.1	Bal31 pTF729
pTF833	1–323[b]	833–1182	77.1	XhoI pAX9m
pTdl835	1–42	835–1182	44.7	Bal31 PTF729
pTdl878	1–42	878–1182	39.8	Bal31 PTF729
pTF893	1–323[b]	893–1182	70.2	BspMI pJ2
pTdl919	1–42	919–1182	35.1	Bal31 pTF893
pTdl920	1–42	920–1182	35.0	Bal31 pTF893
pTdl925	1–42	925–1182	34.4	Bal31 pTF893
pTdl935	1–42	935–1182	33.2	Bal31 pTF893
pTdl938	1–42	938–1182	32.9	Bal31 pTF893

[a] Numbered according to Yanofsky *et al.* (1981)
[b] The pTF constructs contain from 2 to 7 amino acids at the *trpE*-v-*fps* junction sequence encoded by the pUC and pATH poly linker regions
[c] Numbered according to Shibuya & Hanafusa (1982)
[d] Calculated molecular weight in kDa based on amino acid sequences
[e] Parent plasmid for construction and enzyme used to generate the 5' end of v-*fps* coding sequences
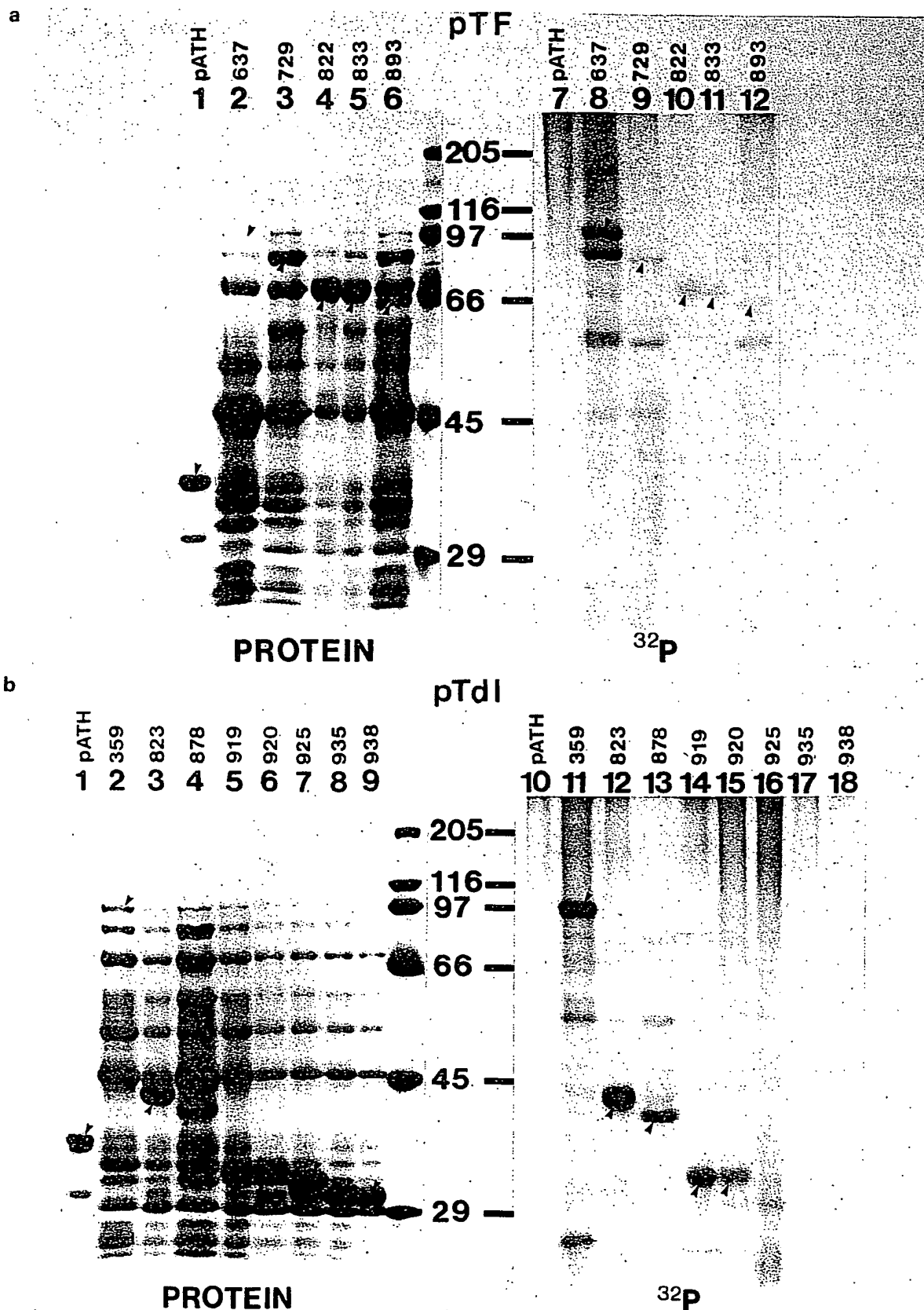
Figure 2   Expression and phosphorylation of *trpE*-v-*fps* fusion proteins in *E. coli*. a: pTF plasmids. *E. coli* containing pATH-1 (lanes 1 and 7), pTF637 (lanes 2 and 8), pTF729 (lanes 3 and 9), pTF822 (lanes 4 and 10), pTF833 (lanes 5 and 11) and pTF893 (lanes 6 and 12) plasmids were grown in suspension culture and the tryptophan operon was induced for 2 h. The induced cells were metabolically labeled with $^{32}P_i$, lysed in SDS–urea buffer, and whole cell lysates analyzed by electrophoresis through 10% SDS–polyacrylamide gels (see Materials and methods). Proteins were detected by staining with Coomassie blue (lanes 1–6); phosphoproteins were identified by autoradiography of the same gel for 4 h (lanes 7–12). b: pTdl plasmids. *E. coli* containing pATH-1 (lanes 1 and 10), pTd1359 (lanes 2 and 11), pTd1823 (lanes 3 and 12), pTd1878 (lanes 4 and 13), pTd1919 (lanes 5 and 14), pTd1920 (lanes 6 and 15), pTd1925 (lanes 7 and 16), pTd1935 (lanes 8 and 17) and pTd1938 (lanes 9 and 18) were grown, induced and labeled with $^{32}P_i$ as above. Proteins were analyzed by electrophoresis and staining with Coomasie blue (lanes 1–9), followed by autoradiography for 4 h (lanes 10–18). The mobilities of size markers and their molecular weight ($\times 10^3$) are indicated. The positions of the 37-kDa pATH protein and of the *trpE*-v-*fps* fusion proteins are indicated by arrows

Table 2   Kinase activity of trpE-v-fps polypeptides

| trpE-v-fps protein | Kinase activity[a] | | | | (% Soluble Protein) |
| | Autophosphorylation | | Phosphorylation | | |
	in vivo[b]	in vitro[c]	enolase[d]	poly(glu,tyr)[e]	
pATH	0	0	0	0	73
pTd1359	16	62	68	722	20
PTd1823	389	340	280	3200	1
pTd1878	218	200	231	7308	6
pTd1919	12	3	4	60	11
pTd1920	5	0	0	35	18
pTd1925	0	0	0	0	12
pTd1935	0	0	0	0	11
pTd1938	0	0	0	0	10
pTF637	14	27	36	762	42
pTF729	8	24	70	1446	20
pTF822	25	93	91	1304	7
pTF833	17	35	37	877	10
PTF893	5	30	124	1498	44

[a] Kinase activity is measured as c.p.m. of ^{32}P ($\times 10^3$) incorporated per 40 pmole soluble trpE-v-fps protein
[b] ^{32}P metabolically incorporated into trpE-v-fps fusion protein
[c] Autophosphorylation of trpE-v-fps proteins in crude bacterial lysates
[d] Phosphorylation of acid-denatured enolase by trpE-v-fps fusion proteins in crude bacterial lysates
[e] Phosphorylation of poly(glu,tyr) (80:20) by crude bacterial lysates

proteins of the predicted sizes (Figure 2a, lanes 1–6; and Figure 2b, lanes 1–9). Only the product of pTF49 with a molecular weight of 167 kDa did not accumulate to significant levels (data not shown). In some cases the identities of these novel polypeptides were confirmed by their specific immunoprecipitation from lysates of [^{35}S]methionine-labeled E. coli with anti-fps rat antiserum against P130$^{gag-fps}$ (data not shown). The trpE-v-fps proteins expressed from these plasmids represented up to 15–20% of total cellular protein synthesized during the induction period. The percentage of trpE-v-fps protein which was soluble following bacterial cell lysis varied widely from 1% (for pTd1823) to 44% (for pTF893) as detailed in Table 2.

Tyrosine phosphorylation of trpE-v-fps *fusion proteins in bacteria defines a minimal catalytic domain required for autophosphorylation*

Following induction of the tryptophan operon, E. coli containing the pATH or trpE-v-fps expression plasmids were metabolically labeled with ^{32}P$_i$ for 20 min, and lysates of the radiolabeled cells were analyzed by electrophoresis (Figure 2). Autoradiography of the stained gels showed that all of the pTF trpE-v-fps fusion proteins incorporated ^{32}P (Figure 2a, lanes 8–12), unlike the parental 37-kDa trpE protein (Figure 2a, lane 7).

Phosphoamino acid analysis of these labeled pTF trpE-v-fps proteins showed that they were phosphorylated exclusively at tyrosine (Figure 3). Furthermore, tryptic phosphopeptide analysis revealed that they were phosphorylated at the same tyrosine residues in E. coli as was P130$^{gag-fps}$ in rat-2 cells (see below). These results suggested that these bacterial trpE-v-fps proteins were enzymatically active tyrosine kinases capable of authentic autophosphorylation. The smallest pTF protein (pTF893) contains residues 893–1182 of P130$^{gag-fps}$, which are clearly capable of catalyzing autophosphorylation.

The span of pTdl hybrid proteins is considerably greater than that of the pTF products (see Table 1). The largest (pTd1359) contains virtually the entire v-fps coding sequence including all FSV coding elements required for neoplastic transformation of chicken embryo fibroblasts (Foster & Hanafusa, 1984), whereas the smallest (pTd1938) lacks a significant part of the highly conserved sequence common to all PTKs. To assay the ability of the pTdl proteins to autophosphorylate in vivo, induced cells were labeled with ^{32}P$_i$ and whole cell lysates were analyzed by electrophoresis and autoradiography (Figure 2b, lanes 11–18). ^{32}P incorporation was detected for pTd1920 and all larger proteins (Figure 2b, lanes 11–15), suggesting that N-terminal v-fps sequences could be deleted to P130$^{gag-fps}$ amino acid 920 while still retaining autophosphorylating catalytic activity. The only prominent phosphoproteins in these cells were the pTdl trpE-v-fps polypeptides themselves. No phosphorylation was detected for pTd1925–938 (lanes 16–18) or for the parental pATH protein (lane 10), indicating that the ability of v-fps polypeptides to autophosphorylate in bacteria was entirely lost with the deletion of the five amino acids between P130$^{gag-fps}$ residues 920 and 925.

Quantitation of the in vivo phosphorylation of trpE-v-fps proteins (Table 2) indicated that the C-terminal v-fps protein fragments initiating at residues 919 and 920 had significantly lower autophosphorylating activity in vivo than the pTd1878 product with 41 additional v-fps amino acids.

The minimum v-fps *catalytic domain phosphorylates exogenous substrates*

We investigated to what extent the deletion of N-terminal sequences would affect the ability of v-fps fragments to phosphorylate exogenously added substrates in vitro. A sensitive and specific substrate for assaying tyrosine kinase activity in crude cell lysates is a random polymer of glutamic acid and tyrosine (Schieven et al., 1986). Lysates of induced E. coli cultures were incubated with poly(glu,tyr) in the presence of [γ-^{32}P]ATP. The polymer was separated from the reaction mixture on non-sodium dodecyl sulfate (SDS) denaturing polyacrylamide gels.

pTF 637

S

T

Y

pTF 729

S

T

Y

pTF 822

S

T

Y

pTF 833

S

T

Y

Figure 3 Phosphoamino acid analysis of *trpE*-v-*fps* fusion proteins. *trpE*-v-*fps* fusion proteins labeled *in vivo* with ^{32}P were recovered from polyacrylamide gels, hydrolyzed in 6 N HCl, and analyzed for phosphoamino acid content by two-dimensional separation using electrophoresis at pH 1.9 and 3.5. The migration of phosphoamino acid markers, revealed by ninhydrin staining, is indicated for phosphoserine (S), phosphothreonine (T) and phosphotyrosine (Y)

After extensive washing to remove unincorporated label, the gels were dried and autoradiographed. As shown in Figure 4 and quantitated in Table 2, the abilities of the pTF and pTdl proteins to phosphorylate poly(glu,tyr) roughly correlated with autophosphorylation *in vivo*. pTdl919 and 920 both phosphorylated the substrate, while pTdl925 had no activity. Thus, enzymatic function was again abolished by the loss of v-*fps* residues 920–925. The activity of pTdl920 in poly(glu,tyr) phosphorylation was slightly lower than that of pTdl919, which in turn was markedly less active than pTdl878 (Table 2). Enolase is a physiological substrate for P130$^{gag\text{-}fps}$ in FSV-transformed fibroblasts, and we therefore determined the capacity of the bacterial *trpE*-v-*fps* polypeptides to phosphorylate enolase *in vitro*. Acid-denatured enolase (5 μg) was incubated with crude lysates of induced *E. coli* cultures in the presence of [γ-^{32}P]ATP. Enolase was specifically phosphorylated by all of the pTF *trpE*-v-*fps* proteins (Figure 5a). Indeed in each case only enolase and the pTF *trpE*-v-*fps* product were strongly phosphorylated, despite the presence of normal bacterial proteins. All of the pTdl proteins larger than pTdl919 inclusive phosphorylated

enolase (Figure 5b). The *in vitro* kinase activity of the pTdl919 protein measured by enolase phosphorylation was considerably lower than that of the pTdl878 product. Longer exposure of the autoradiograph of Figure 5b showed that pTdl919 autophosphorylated *in vitro* as well as phosphorylating enolase. No *in vitro* kinase activity assayed by enolase phosphorylation and autophosphorylation was detected for pTdl920–938 (Figure 5b, lanes 13–16) or for the 37-kDa pATH protein (Figure 5a, lane 7).

The specific activities of the various *trpE*-v-*fps* proteins for phosphorylation of enolase and poly(glu,tyr) were quantitated following correction for inactive insoluble *trpE*-v-*fps* protein, and were compared with the values obtained for autophosphorylation *in vitro* and *in vivo* (Table 2). The activities of the pTF proteins were quite similar for all parameters, differing at most by 3- to 4-fold between pTF637 and pTF893. The activities of the pTdl *trpE*-v-*fps* proteins showed more variation. pTdl823 and pTdl878 were about 5-fold more active than pTdl359, while the activity of pTdl919 was 50- to 100-fold less than that of pTdl878. pTdl920 was less active again than
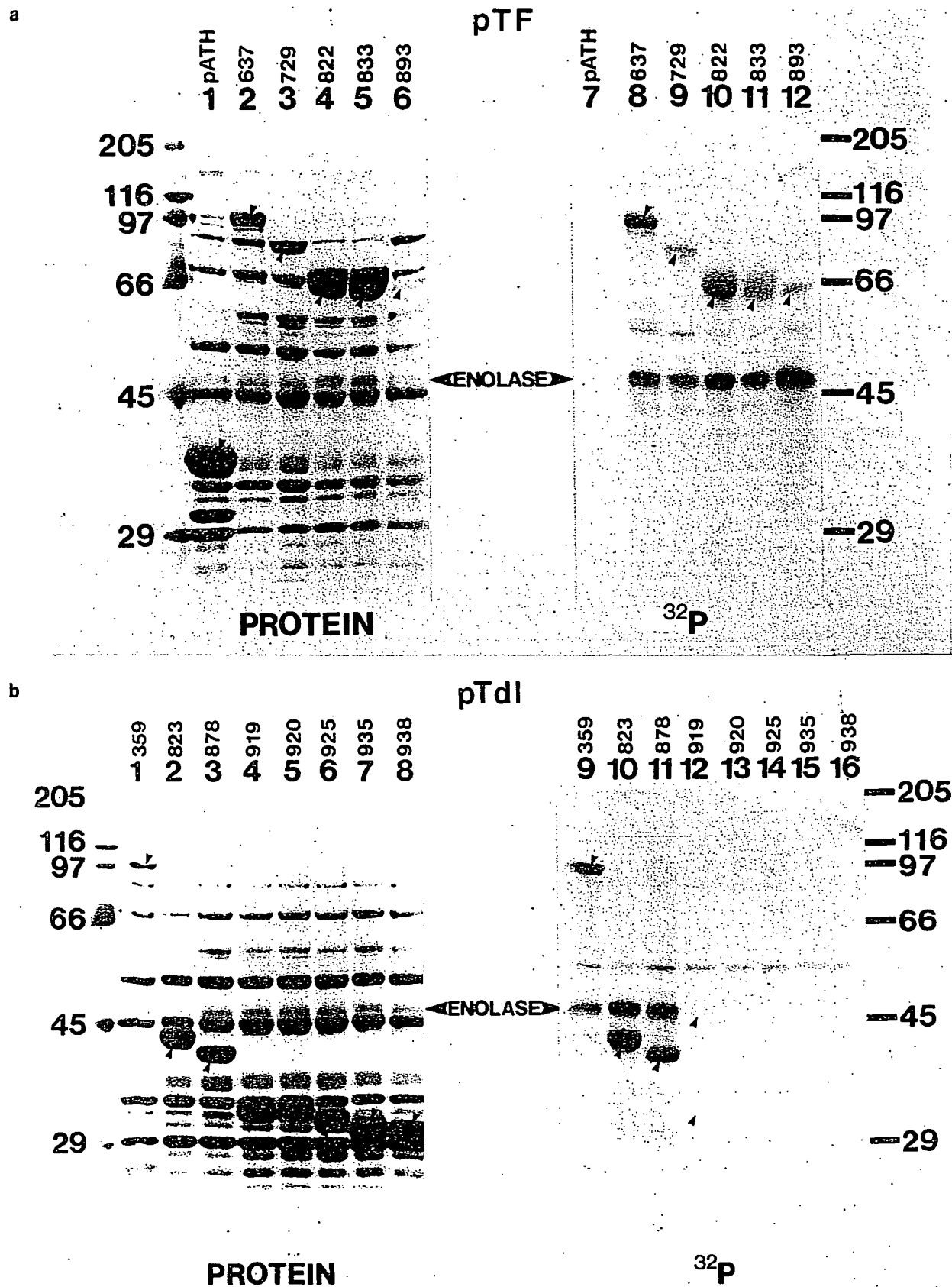
Figure 5   *In vitro* autophosphorylation and phosphorylation of enolase by pTF (a) or PTdl (b) bacterial *trpE*-v-*fps* fusion proteins. Lysates were made from induced *E. coli* containing the following plasmids. a: pATH (lanes 1 and 7), pTF637 (lanes 2 and 8), pTF729 (lanes 3 and 9), pTF822 (lanes 4 and 10), pTF833 (lanes 5 and 11), pTF893 (lanes 6 and 12). b: pTd1359 (lanes 1 and 9), pTd1823 (lanes 2 and 10), pTd1878 (lanes 3 and 11), pTd1919 (lanes 4 and 12), pTd1920 (lanes 5 and 13), pTd1925 (lanes 6 and 14), pTd1935 (lanes 7 and 15), and pTd1938 (lanes 8 and 16). Lysates were incubated with 5 μg acid denatured enolase in the presence of 5 μCi [γ-^{32}P]ATP for 15 min at 37°C. Reactions were analyzed by electrophoresis on 10% SDS−polyacrylamide gels, staining with Coomassie blue (a, lanes 1−6; b, lanes 1−8) and autoradiography for 4 h (a, lanes 7−12; b, lanes 9−16). The positions of the pATH and *trpE*-v-*fps* proteins, and of enolase are indicated with arrows
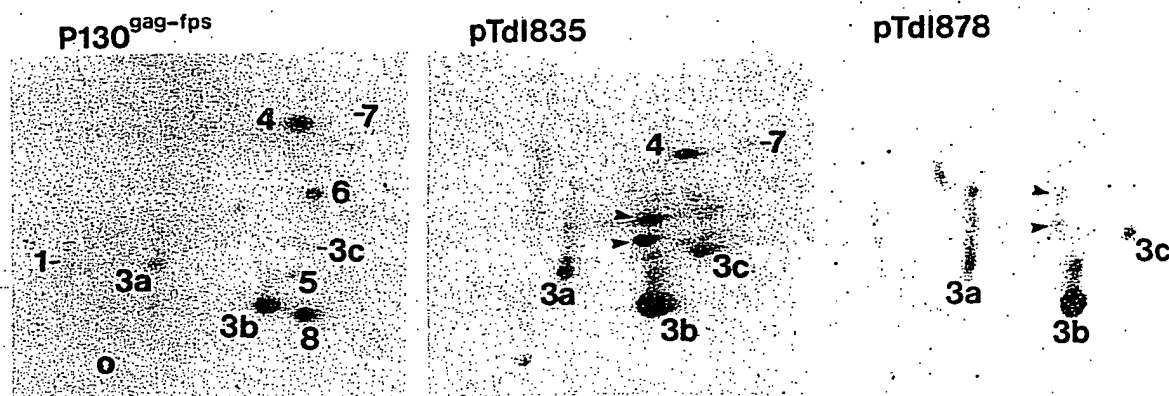
Figure 6 Comparative tryptic phosphopeptide analysis of *trpE*-v-*fps* bacterial fusion proteins and Fujinama avian sarcoma virus (FSV) P130$^{gag-fps}$. Two-dimensional tryptic peptide analysis was undertaken as follows: P130$^{gag-fps}$, P130$^{gag-fps}$ isolated from ^{32}P-labeled FSV-transformed rat-2 cells; pTdl835, pTdl835 P44$^{trpE-v-fps}$ protein from ^{32}P-labeled *E. coli;* pTdl878, pTdl878 P40$^{trpE-v-fps}$ ^{32}P-labeled bacterial protein. Gel-purified proteins were oxidized, digested with trypsin and separated in two dimensions. Electrophoresis at pH 2.1 was from left to right and chromatography in *N*-butanol:acetic acid:water:pyridine (75:15:60:50 by volume) was from bottom to top. Peptides 1, 3a–3c, 4 and 7 of FSV P130$^{gag-fps}$ are known to contain phosphotyrosine, while peptides 5, 6, and 8 contain phosphoserine (Weinmaster *et al.*, 1985, 1984). Arrows indicate phosphorylated peptides specific to the *trpE*-v-*fps* fusion proteins in *E. coli*
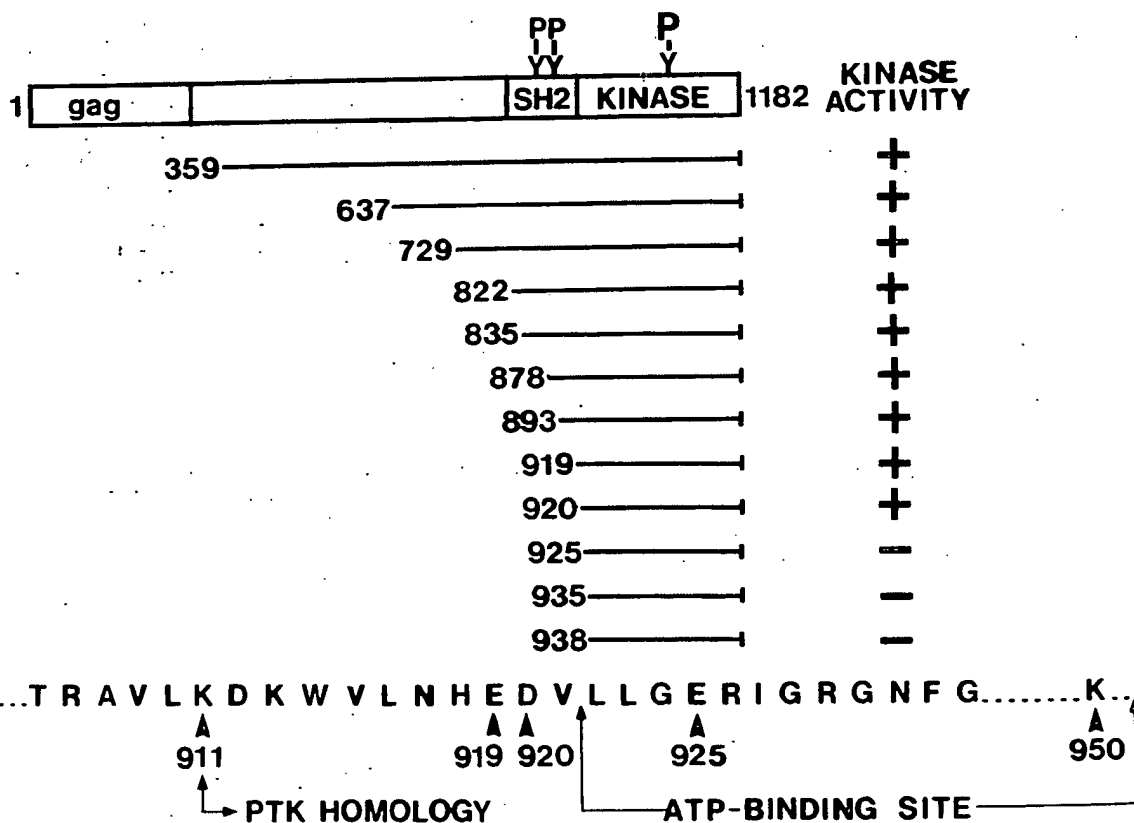
Figure 7 Schematic summary of v-*fps* domain structure indicated by *trpE*-v-*fps* bacterial expression vectors. The structure of Fujinama avian sarcoma virus P130$^{gag-fps}$ at top shows the kinase catalytic and SH2 regions. Confirmed and predicted sites of tyrosine phosphorylation are given. The N-terminal border of the kinase domain is deduced from the catalytic activities of v-*fps* polypeptide fragments expressed in *E. coli*, as shown. At bottom, a partial amino acid sequence of P130$^{gag-fps}$ residues 906–950 is given. The start of homology of P130$^{gag-fps}$ with all protein-tyrosine kinases (PTK) (Hunter & Cooper, 1985; Sadowski *et al.*, 1986), the positions of amino acids 919, 920 and 925, as well as the previously predicted borders of the ATP-binding site (Weinmaster *et al.*, 1986) are indicated

interact with the γ-phosphate of bound ATP (Kamps *et al.*, 1984; Weinmaster *et al.*, 1986). Substitution of this residue with arginine by oligonucleotide-directed mutagenesis has shown that lys-950 is critical for P130$^{gag-fps}$ kinase activity, and may participate directly in catalyzing phosphotransfer at the active site (Weinmaster *et al.*, 1986). In addition, residues 922–950 show primary

structural similarity to the adenine nucleotide-binding sites of NAD-binding dehydrogenases such as glyceraldehyde 3-phosphate dehydrogenase and the AMP-binding site of adenylate kinase for which the complete tertiary structures have been solved (Buehner *et al.*, 1974; Shulz *et al.*, 1974; Weinmaster *et al.*, 1986). The nucleotide-binding folds of these proteins assume the conformation β1

strand–loop–α helix–β2 strand, forming a hydrophobic pocket for the adenine group from which the ribose and phosphates protrude (Buehner et al., 1974; Pai et al., 1977). Given the collateral evidence that this is indeed the protein kinase ATP-binding site, it seems likely that residues 922–950 of P130$^{gag-fps}$ adopt a similar conformational structure (Sternberg & Taylor, 1984). In this case, residues 922–928 of P130$^{gag-fps}$ would form a β strand, while the following highly conserved glycine-rich sequence (residues 928–933 of P130$^{gag-fps}$) would form a loop joining this β strand to an α helix. Deletion of v-*fps* sequences to amino acid 925 would remove most of the predicted β1 strand of the ATP-binding site and might thereby abolish catalytic function by interfering with ATP binding.

Since all the pTd1 bacterial proteins had 42 N-terminal *trpE* amino acids, it might be argued that the loss of catalytic function in the pTd1925, 935 and 938 products resulted not from deletion of sequences required for catalysis but from conformational distortion of the catalytic domain induced by proximity of the short *trpE* sequence. However, the precise correspondence observed between complete loss of kinase activity and encroachment on the predicted ATP-binding site suggest that this assay is a reliable indicator of v-*fps* elements absolutely required for catalytic activity.

The relatively small, but detectable, levels of kinase activity exhibited by the pTd1919 and 920 proteins suggests that N-terminal v-*fps* sequences can be deleted to the start of the ATP-binding site without preventing ATP- or substrate-binding or the transfer of phosphate from one to the other. Sequences N-terminal to residue 919 are therefore not obligatory for binding and phosphorylation of substrates such as enolase and poly(glu,tyr). Several observations, however, suggest that sequences within the non-catalytic domain immediately N-terminal to the core catalytic region formed by P130$^{gag-fps}$ residues 920–1182 may interact with this region and influence kinase function.

Firstly, the additional 41 residues in the pTd1878 protein stimulate activity by about 100-fold compared with the pTd1919 protein. These sequences may, therefore, contribute to some aspect of catalytic function, or may stabilize the structure of the catalytic domain. Secondly, in analyzing the phosphopeptides of the bacterial *trpE*-v-*fps* proteins we observed that autophosphorylation at sites represented by tryptic phosphopeptides 4 and 7 was lost in the pTd1878 protein but was present in the pTd1835 product. The simplest explanation for these data is that these sites are located between P130$^{gag-fps}$ residues 835 and 878 within the SH2 domain, although the results could also be interpreted as an indirect effect of the deletion. In either event this observation implies an interaction between the SH2 domain and the core catalytic region. We have suggested that the SH2 domain of P130$^{gag-fps}$ is involved in the recognition of cellular proteins in vertebrate cells (Sadowski et al., 1986). The corresponding region of Rous sarcoma virus p60^{v-src} has been implicated in defining substrate specificity (Jove et al., 1986). It is possible that non-catalytic v-*fps* domains such as the adjacent SH2 region may be required for the recognition of specific targets involved in neoplastic transformation of vertebrate cells.

The phosphorylation of exogenous substrates by *trpE*-v-*fps* bacterial proteins was rather restricted. The predominant phosphoproteins of labeled bacteria and bacterial lysates were autophosphorylated *trpE*-v-*fps* proteins or exogenously added enolase, despite the presence of numerous bacterial proteins (Figure 5a, lanes 1–6; Figure 5b, lanes 1–8). The only *E. coli* protein to be consistently phosphorylated was a 165-kDa polypeptide (Figure 2a, lanes 8 and 9; Figure 5a, lanes 8 and 9) that contains phosphotyrosine in induced cells (data not shown). In this regard, bacterial v-*fps* polypeptides behaved differently from v-*abl* proteins expressed in *E. coli*, which phosphorylate numerous bacterial proteins at tyrosine (Wang et al., 1982; Prywes et al., 1985).

In summary, our data suggest that the start of the v-*fps* tyrosine kinase catalytic domain is effectively defined by the ATP-binding site and that the kinase domain can function as an autonomous unit. The C-terminal 263 residues of P130$^{gag-fps}$ apparently possess all the information required for substrate binding and phosphotransfer, although additional N-terminal sequences may be required for full catalytic activity. Previous work of Yaciuk & Shalloway (1986) has suggested that the C-terminus of the *src* family kinase domain is defined by a conserved leucine (residue 1173 of P130$^{gag-fps}$) that is required for protein stability in vertebrate cells.

A major obstacle to the biochemical and structural analysis of retroviral transforming proteins such as v-*fps* is a lack of material. By the expression of relatively soluble v-*fps* polypeptides in *E. coli* we have acquired a convenient system for the study of v-*fps* kinase activity. The bacterial v-*fps* proteins are expressed to high levels and autophosphorylate at physiological sites. We expect that the vectors described here will enable us to thoroughly study specific regions of the kinase domain by combined mutational, biochemical and structural analysis.

Materials and methods

Bacteria and plasmids

The RR1 strain of *E. coli* (hsd20, ara14, proA2, lacY1, galk2, rpsL20, xyl15, mtl1, supE44) was used for most transformations (Bolivar et al., 1977). The construction of the plasmid pJ2, and of the mutants RX18m, RX15m and AX9m derived from pJ2 have been described previously (Stone et al., 1984; Sadowski et al., 1986). FSV DNA sequences used in these constructs derive from a molecular clone isolated by Shibuya & Hanafusa (1982).

Construction of bacterial expression plasmids

Figure 1 summarizes the construction of the plasmids used in these experiments. To generate v-*fps* encoding fragments of decreasing size we cut the FSV genome at unique restriction enzyme sites and at unique sites created by linker insertion mutagenesis. Digestion of wild type FSV with SmaI, PvuII/SmaI or BspMI/SmaI generated v-*fps* encoding fragments starting at amino acid 49, 729 and 893, respectively (Figure 1, panels a and b). The BspMI/SmaI fragment was made blunt by end-filling with Klenow enzyme. Digestion of the FSV XhoI linker insertion mutants RX18m, RX15m and AX9m with XhoI/HindIII generated fragments encoding v-*fps* from amino acids 637, 822 and 833 (panel c). Blunt ended fragments were cloned into the SmaI site of pUC vectors while the XhoI/HindIII fragments were cloned into the SalI/HindIII sites of pUC8. From these constructs, the EcoRI/HindIII inserts were subcloned into the proper reading frame of pATH-1, pATH-3 or pATH-11 (Figure 1, panels a–c). A preliminary description of the pTF plasmids has been given elsewhere (Sadowski et al., 1986).

To construct pTd1823, pTd1835 and pTd1878, the pTF729 plasmid was digested with KpnI which cleaves at the codon for amino acid 821 (Figure 1, bottom of panel a). Linearized DNA was treated with Bal31 nuclease (New England Biolabs) for 2–10 min at 30°C. Bal31-digested DNA was cut with EcoRV which cleaves between the codons for amino acid 42 and 43 of trpE. The fragments were purified by electrophoresis on 0.7% agarose. Following end-filling with Klenow and religation, the plasmids were retransformed into RR1. Clones were screened by analysis of protein products on 10% polyacrylamide gels. Those expressing novel proteins of the proper size were sequenced from double-stranded DNA using the oligonucleotide primer 5′ GAACAAAATTAGAGAATA 3′ which corresponds to nucleotide sequence 142–159 of the E. coli tryptophan operon (Yanofsky et al., 1981). The construction of pTd1919, 920, 925 and 935 was as described above except that pTF893 was linearized with KpnI which cuts in the polylinker region between the coding sequence for trpE and v-fps (Figure 1, panel d). Bal31 reactions were for 20 or 30 min using Slow Bal31 (International Biotechnologies Inc.). pTd1938 was constructed by digestion of pTF729 with NotI and EcoRV. The NotI 5′ protrusion was filled in with Klenow and the vector religated. The pTd1359 plasmid was constructed by digesting pTF49 with EcoRV and NcoI, limited S_1 nuclease treatment and religation.

Growth of bacteria and induction of protein expression

Cells containing expression plasmids were grown overnight at 37°C in 1 ml M9 medium supplemented with 0.5% w/vol casamino acids, 10 µg/ml thiamine, 20 µg/ml tryptophan and 50 µg/ml ampicillin. The overnight culture was diluted 1/10 into 1 ml fresh M9 supplemented with casamino acids, thiamine and ampicillin and grown for 1 h with strong aeration at 37°C. 5 µl of 1 mg/ml indole acrylic acid (Sigma) were added in ethanol to induce the tryptophan operon. The cells were grown for a further 2 h and harvested by centrifugation.

To analyze whole cell protein, the cells were resuspended in 25 µl SDS–urea buffer (10 mM sodium phosphate pH 7.2, 1% vol/vol 2-mercaptoethanol, 1% w/vol SDS, 6 M urea). The lysates were incubated at 37°C for 30 min and 25 µl of 2 × SDS buffer (20% vol/vol glycerol, 10% vol/vol 2-mercaptoethanol, 4.6% w/vol SDS, 0.125 M Tris-HCl pH 6.8) was then added. The samples were heated at 100°C for 3 min and 15-µl aliquots were analyzed by electrophoresis on a 7.5% SDS–polyacrylamide gel as described (Weinmaster et al., 1983).

Metabolic labeling of bacterial and mammalian cells

Cells containing plasmid vectors were grown as described above. Following 2 h induction with indole acrylic acid, the cells were harvested by centrifugation and washed twice in 50 mM Tris-HCl pH 7.5, 10 mM MgCl₂, 150 mM NaCl and resuspended in 50 µl of the same buffer. The suspensions were incubated with 60 µCi $^{32}P_i$ for 20 min at room temperature. The cells were then washed 3 × in TEN buffer (50 mM Tris-HCl pH 7.5, 0.5 mM EDTA, 0.3 M NaCl) and resuspended in 25 µl SDS–urea buffer. The whole cell lysates were analyzed by electrophoresis through 7.5% SDS–polyacrylamide gels.

Growth of FSV-transformed rat-2 cells and metabolic labeling with $^{32}P_i$ were as described (Weinmaster et al., 1983, 1984). P130gag-fps was immunoprecipitated from radiolabeled cells with R254E anti-p19gag monoclonal antibody (Ingman-Baker et al., 1984) and purified by electrophoresis through 7.5% SDS–polyacrylamide gels.

Protein analysis

Electrophoretic separation of proteins on SDS–polyacrylamide gels, isolation of radiolabeled proteins from gels, and subsequent phosphoamino acid analysis and tryptic phosphopep-

tide mapping were carried out exactly as described previously (Weinmaster et al., 1983, 1984).

Preparation and assay of bacterial lysates for kinase activity

Bacteria were grown and induced as described above. The pellets from 1-ml cultures were washed once with TEN buffer and resuspended in 10 µl E. coli lysis buffer (50 mM Tris-HCl pH 7.5, 0.3 M NaCl, 1% vol/vol NP40, 20 mM MgCl₂). 10 µl E. coli lysis buffer plus 2 mg/ml lysozyme were added and the suspension was incubated on ice for 30 min. 10 µl of the same buffer plus 0.3 mg/ml DNAaseI was then added and the suspension held on ice for a further 60 min. For assay of enolase phosphorylation, 10 µl of the lysate were mixed with 5 µg acid-denatured rabbit muscle enolase and kinase reaction buffer (100 mM Hepes pH 7.5, 10 mM MnCl₂). 5 µCi [γ-³²P]ATP were added in reaction buffer to bring the total volume to 30 µl. The reactions were incubated at 37°C for 15 min and stopped by the addition of 30 µl 2 × SDS sample buffer. The samples were analyzed by electrophoresis through 10% polyacrylamide gels and autoradiography. For quantitation of trpE-v-fps protein in each reaction, the Coomassie blue-stained gels were densitometrically scanned and the integrals from the protein peaks used to calculate the amount of protein per band.

Phosphorylation of poly(glu,tyr) (80:20) polymer was assayed using crude lysates as described above. 10 µl crude lysate were mixed with 10 µg poly(glu,tyr) (Sigma) in kinase reaction buffer. 2 µCi [γ-³²P]ATP were added in reaction buffer to bring the volume to 30 µl. After 15 min at 37°C the reaction was stopped by the addition of 1 µl 100 mM cold ATP, 2 µl of 500 mM EDTA and 40 µl 2 × sample buffer (4% vol/vol NP40, 10% vol/vol glycerol, 125 mM Tris-HCl pH 6.8, 0.002% bromophenol blue). The reactions were electrophoresed on 5% polyacrylamide gels containing 8 M urea, 2% NP40, 0.325 M Tris-HCl pH 8.8 (Schieven et al., 1986). The gels were fixed and washed extensively in 10% acetic acid, 50 mM sodium pyrophosphate. Gels were dried and autoradiographed prior to excising the poly(glu,tyr) bands and quantitation of ³²P counts. To measure the amount of trpE-v-fps protein in the poly(glu,tyr) reactions, 10 µl of the bacterial lysate were run on a 10% SDS–polyacrylamide gel, and the Coomassie blue-stained gel densitometrically scanned.

Quantitation of solubilities of trpE-v-fps fusion proteins

E. coli were grown and induced as described above, except that 40 µCi [³⁵S]methionine was added at the time of induction. The cells were lysed as above and the lysates centrifuged at 12 000 g for 20 min. The soluble and insoluble fractions were run on 10% polyacrylamide gels. The amount of protein in each fraction was determined by excising the appropriate bands and counting the incorporated [³⁵S]methionine. Counts from background bands were subtracted in calculation of protein levels. The identities of the trpE-v-fps proteins were verified immunologically by immunoprecipitation of [³⁵S]methionine-labeled clarified lysates with 4 µl anti-fps rat antiserum.
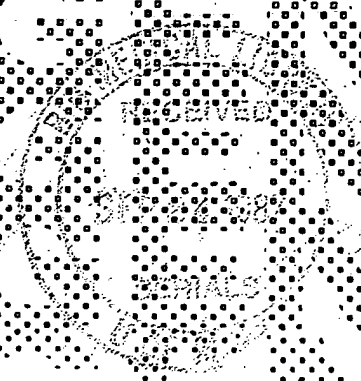
References

Bolivar, F., Rodriguez, R.L., Green, P.I. & 5 others (1977). *Gene*, **2**, 95–99.

Buehner, M., Ford, G.C., Moras, D., Olsen, K.W. & Rossman, M.G. (1974). *J. Mol. Biol.*, **90**, 25–49.

Cooper, J.A., Gould, K.L., Cartwright, C.A. & Hunter, T. (1986). *Science*, **231**, 1431–1434.

Downward, J., Yarden, Y., Mayes, E. & 6 others (1984). *Nature*, **307**, 521–527.

Foster, D. & Hanafusa, H. (1984). *J. Virol.*, **48**, 744–751.

Garber, A., Cross, F. & Hanafusa, H. (1985). *Mol. Cell. Biol.*, **5**, 2781–2788.

Hunter, T. & Cooper, J. (1985). *Annu. Rev. Biochem.*, **54**, 897–930.

Ingman-Baker, J., Hinze, E., Levy, J.G. & Pawson, T. (1984). *J. Virol.*, **50**, 572–578.

Jove, R., Garber, E.A., Iba, H. & Hanafusa, H. (1986). *J. Virol.*, **60**, 849–857.

Kamps, M.P., Taylor, S.A. & Sefton, B.M. (1984). *Nature*, **310**, 589–592.

Levinson, A.D., Courtneidge, S. & Bishop, J.M. (1981). *Proc. Natl. Acad. Sci. USA*, **78**, 1624–1628.

Pai, E.F., Sachsenheimer, W., Schirmer, R.H. & Schultz, G.E. (1977). *J. Mol. Biol.*, **114**, 37–45.

Pellman, D., Garber, E.A., Cross, F.R. & Hanafusa, H. (1985). *Nature*, **314**, 374–377.

Prywes, R., Foulkes, J.G. & Baltimore, D. (1985). *J. Virol.*, **54**, 114–122.

Sadowski, I., Stone, J.C. & Pawson, T. (1986). *Mol. Cell Biol.*, **6**, 4396–4308.

Shulz, G.E., Elzinga, M., Marx, F. & Schirmer, R.H. (1974). *Nature*, **250**, 120–123.

Schieven, G., Thorner, J. & Martin, G.S. (1986). *Science*, **231**, 390–393.

Shibuya, M. & Hanafusa, H. (1982). *Cell*, **30**, 787–795.

Shtivelman, E., Lifshitz, B., Gale, R.P. & Canaani, E. (1985). *Nature*, **315**, 550–554.

Sternberg, M.J.E. & Taylor, W.R. (1984). *FEBS Lett.*, **175**, 387–392.

Stoker, A.W., Enrietto, P.J. & Wyke, J.A. (1984). *Mol. Cell. Biol.*, **4**, 1508–1514.

Stone, J.C., Atkinson, T., Smith, M. & Pawson, T. (1984). *Cell*, **37**, 549–558.

Stone, J.C. & Pawson, T. (1985). *J. Virol.*, **55**, 721–727.

Wang, J.Y.J., Queen, C. & Baltimore, D. (1982). *J. Biol. Chem.*, **257**, 13181–13184.

Weinmaster, G., Hinze, E. & Pawson, T. (1983). *J. Virol.*, **45**, 29–41.

Weinmaster, G., Zoller, M.J., Smith, M., Hinze, E. & Pawson, T. (1984). *Cell*, **37**, 559–568.

Weinmaster, G., Zoller, M.J. & Pawson, T. (1986). *EMBO J.*, **5**, 69–76.

Yaciuk, P. & Shalloway, D. (1986). *Mol. Cell. Biol.*, **6**, 2807–2819.

Yanofsky, C., Platt, T., Crawford, L.P. & 5 others (1981). *Nucleic Acids Res.*, **9**, 6647–6668.

ONCOGENE

An International Journal

Volume 1　　　Number 2　　　May 1987

Exhibit 12

25. Zhang, F., Strand, A., Robbins, D., Cobb, M. H. & Goldsmith, E. Nature **357**, 704–710 (1994).
26. Han, J., Lee, J.-D., Tobias, P. S. & Ulevitch, R. J. J. biol. Chem. **268**, 25009–25014 (1993).
27. Lee, J.-D. et al. Proc. natn. Acad. Sci. U.S.A. **90**, 9930–9934 (1993).
28. Dong, Z., Qi, X. & Fidler, I. J. J. exp. Med. **177**, 1071–1077 (1993).
29. Novogrodsky, A. et al. Science **264**, 1319–1322 (1994).
30. Sung, S.-J., Walters, J. A. & Fu, S. M. J. exp. Med. **176**, 897–901 (1992).
31. Han, J., Brown, T. & Beutler, B. J. exp. Med. **171**, 465–475 (1990).
32. Kaspar, R. L & Gehrke, L. J. Immun. **153**, 277–286 (1994).
33. Kruys, V., Kemmer, K., Shakhov, A., Jongeneel, V. & Beutler, B. Proc. natn. Acad. Sci. U.S.A. **89**, 673–677 (1992).
34. Keffer, J. et al. EMBO J. **10**, 4025–4031 (1991).
35. Zhang, W. et al. Molec. cell. Biol. **13**, 7652–7665 (1993).
36. Huang, L.-Y., Tholanikunnel, B. G., Vakalpoulou, E. & Malbon, C. C. J. biol. Chem. **268**, 25769–25775 (1993).
37. Lin, T.-A. et al. Science **266**, 653–656 (1994).
38. Lee, J. C. et al. Int. J. Immunother. **6**, 1–12 (1990).
39. Grantham, R., Gautier, C., Gouy, M., Jacobzone, M. & Mercier, R. Nucleic Acids Res. **9**, r43–r74 (1981).
40. Livi, G. P. et al. Molec. cell. Biol. **10**, 2678–2686 (1990).
41. Ausubel, F. M. et al. Current Protocols in Molecular Biology (Greene & Wiley, Brooklyn, NY, 1987).
42. Wood, W. I., Gitschier, J., Lasky, L. A. & Lawn, R. M. Proc. natn. Acad. Sci. U.S.A. **82**, 1585–1588 (1985).
43. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. J. molec. Biol. **215**, 403–410 (1990).
44. Hochuli, E., Bannwarth, W., Dobeli, H., Gentz, R. & Schacher, A. Bio/Technology **6**, 1321–1325 (1988).

Crystal structure of the tyrosine kinase domain of the human insulin receptor

Stevan R. Hubbard[*], Lei Wei[†], Leland Ellis[†] & Wayne A. Hendrickson[*‡]

[*] Department of Biochemistry and Molecular Biophysics, and [‡] Howard Hughes Medical Institute, Columbia University, New York, New York 10032, USA
[†] W. M. Keck Center for Genome Informatics, Institute of Biosciences and Technology, Texas A&M University, Houston, Texas 77030, USA

The X-ray crystal structure of the tyrosine kinase domain of the human insulin receptor has been determined by multiwavelength anomalous diffraction phasing and refined to 2.1 Å resolution. The structure reveals the determinants of substrate preference for tyrosine rather than serine or threonine and a novel autoinhibition mechanism whereby one of the tyrosines that is autophosphorylated in response to insulin, Tyr 1,162, is bound in the active site.

INSULIN activates a number of signalling pathways that regulate cellular metabolism and growth (reviewed in ref. 1). The insulin receptor is a transmembrane glycoprotein[2,3] and a member of the receptor tyrosine kinase family which includes, among others, the receptors for epidermal and platelet-derived growth factors (EGF and PDGF). The extracellular portion of receptors in this family contains the binding site for its particular protein ligand and the tyrosine kinase activity resides in the cytoplasmic portion. In contrast to the EGF and PDGF receptors, which are monomeric and dimerize upon ligand binding (reviewed in ref. 4), the insulin receptor is a disulphide-linked $\alpha_2\beta_2$ heterotetramer. Binding of insulin to the extracellular α-chains is thought to cause a change within the quaternary structure of the receptor that results in autophosphorylation of specific tyrosines in the cytoplasmic portion of the β-chains: two in the juxtamembrane region (~30 residues C-terminal to the transmembrane helix), three in the tyrosine kinase domain (~300 residues), and two in the C-terminal tail (~70 residues)[5–8]. The nature of the autophosphorylation events, whether cis (within a β-chain) or trans (between β-chains), has been the subject of debate[9–12].

Autophosphorylated tyrosines of the EGF and PDGF receptors serve as binding sites for proteins that contain Src-homology-2 (SH2) domains[13], whereas the phosphotyrosines of an insulin-receptor substrate, IRS-1, rather than those of the receptor itself, are the predominant targets for SH2-containing proteins[14]. Although the roles of the phosphotyrosines in the juxtamembrane region and the C-terminal tail have not been fully elucidated, it is well established that autophosphorylation of the three tyrosines in the kinase domain stimulates kinase activity towards exogenous substrates[7]. The tyrosine kinase activity of the insulin receptor has been shown through kinase-inactivating point mutations to be essential for insulin signal transduction[15,16]. A number of nonsense and missense mutations

in the tyrosine kinase region of the gene encoding the insulin receptor have been identified in patients afflicted with non-insulin-dependent diabetes mellitus (NIDDM) (reviewed in ref. 17).

We have used the multiwavelength anomalous diffraction (MAD) phasing method[18] to determine the crystal structure of the unphosphorylated, apo form of the tyrosine kinase domain of the insulin receptor (IRK). The structures of several protein serine/threonine kinases (PSKs) have been reported: cyclic AMP-dependent protein kinase (cAPK)[19], cyclin-dependent kinase 2 (CDK2)[20], mitogen-activated protein kinase (MAPK)[21], and twitchin kinase[22]. The IRK structure reveals those features that are characteristic of members of the protein tyrosine kinase (PTK) family.

Structure determination

A baculovirus/insect cell expression system was used to produce a 306-residue fragment of the β-chain of the human insulin receptor which contains tyrosine kinase activity and the autophosphorylation sites Tyr 1,158, Tyr 1,162 and Tyr 1,163 (L.W., S.R.H., W.A.H. and L.E., manuscript submitted; numbering is according to ref. 3). The N and C termini of the expressed protein were chosen on the basis of proteolytic studies performed on the entire cytoplasmic portion (M_r 48K) of the β-chain. The purified protein was not tyrosine-phosphorylated as judged by western blotting with anti-phosphotyrosine antibody and autophosphorylation experiments demonstrating phosphorylation at three sites upon addition of Mg–ATP.

Attempts to solve the structure of the unphosphorylated, apo form of IRK by molecular replacement using the cAPK structure as a search model were unsuccessful. The structure was solved by the MAD phasing method using a crystal derivatized with ethylmercuric phosphate, which binds to two cysteine thiols. Synchrotron diffraction data were collected at three X-ray wavelengths near the mercury L_{III} absorption edge. The initial MAD
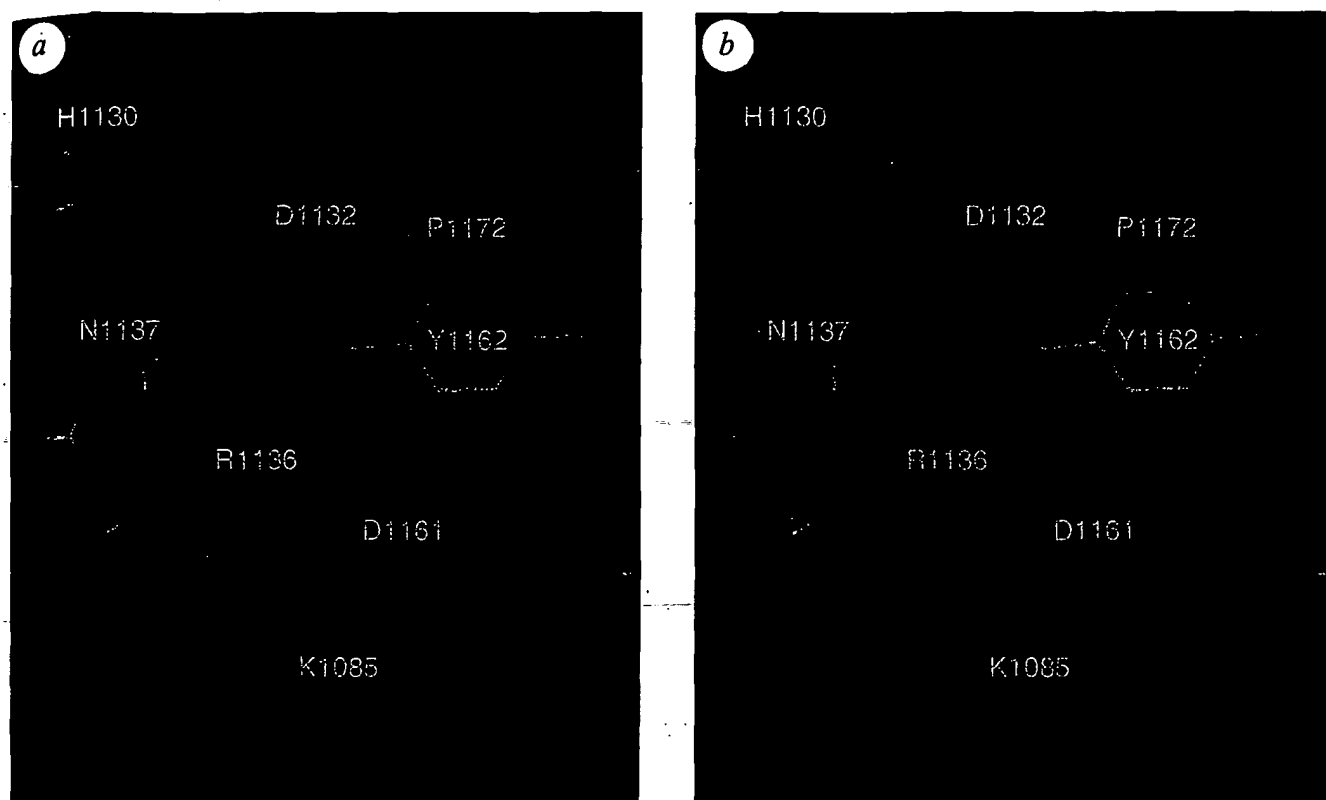
FIG. 1 *a*, The experimental, MAD-phased electron density map at 2.5 Å resolution in the region of the IRK active site, contoured at 1σ. Superposed is the refined atomic model. *b*, The corresponding 2F₀−F꜀ electron density map at 2.1 Å resolution, contoured at 1σ. Figures prepared with SETOR[41].

derived phases to 2.5 Å Bragg spacings were subsequently improved through multiple cycles of model building, refinement and partial model phase combination. The structure has been refined to 2.1 Å resolution with a crystallographic *R*-value of 19.6%. The atomic model includes all but the three N-terminal residues. Electron density maps computed with the MAD-derived phases and with the phases calculated from the atomic model are shown in Fig. 1. Details of crystallization, data collection and analysis, model building and structure refinement are given in Table 1.

Overall topology

IRK is composed of two lobes with a single connection between them, similar to the kinase cores of the PSKs whose structures have been determined[19-22] (Fig. 2a). The N-terminal lobe comprises a twisted β-sheet of five antiparallel β-strands (β1–β5) and one α-helix (αC). The secondary structure nomenclature follows that for cAPK[19]; a structure-based sequence alignment is shown in Fig. 3. IRK lacks the short α-helix B which immediately precedes α-helix C in cAPK[19], as do the other PSKs with known structure. The β-sheet cores of IRK and cAPK are very similar, with a root-mean-square (r.m.s.) deviation of only 0.8 Å for a superposition of 30 Cα atoms in the five β-strands. Differences in the N-terminal lobe are found in the loops between β-strands, including a five-residue insertion between β2 and β3 in IRK, and near the N terminus of αC (Fig. 2c). Although the first eight residues in the IRK structure (981–988) are relatively mobile (main-chain *B*-values >45 Å²), the backbone conformation is extended rather than α-helical; the latter was predicted for PTKs possessing the semiconserved Trp(989)-Glu sequence[23].

The larger C-terminal lobe comprises eight α-helices (αD, αE, αEF, αF-αJ) and four β-strands (β7, β8, β10, β11). The one-and-a-half-turn αEF is only a one-turn helix in cAPK and was

not identified as such in the cAPK structure report. IRK lacks β-strands 6 and 9 of cAPK, although the IRK residues that correspond to β-strand 6 follow a similar course. A superposition of 66 Cα atoms in helices αD–αH of IRK and cAPK gives an r.m.s. deviation of 1.3 Å. In the C-terminal lobe, IRK and cAPK are most dissimilar in the activation loop, between β8 and αEF, and in the connection between αD and αE. The activation loop contains the three tyrosine autophosphorylation sites in IRK and the essential phosphorylation site Thr 197 in cAPK. In the apo IRK structure, the activation loop (five residues longer than in cAPK) contains the two short β-strands β10 and β11 and traverses the cleft between the N- and C-terminal lobes (Fig. 2d). The paths of the apo IRK and cAPK activation loops diverge at Gly 1,149 (Thr 183) and reconverge at Pro 1,172 (Thr 201). Residues 1,152–1,157 in the IRK activation loop are relatively mobile, with main-chain *B*-values exceeding 45 Å². The loop connecting αD and αE, known as the kinase insert region (~100-residue insertion in the PDGF receptor), is ten residues longer in IRK than in cAPK and contains the sequence Pro(1,099)-Gly-Arg-Pro-Pro-Pro, which is reminiscent of proline-rich sequences to which Src-homology-3 (SH3) domains bind[24]. The three consecutive prolines in the IRK structure adopt a left-handed polyproline type II helical conformation, as do the proline-rich peptides in the SH3–peptide complexes whose structures have been determined[25,26]. The proline-rich sequence in IRK is shorter by one helical turn than the canonical SH3-binding sequence; as yet, no IRK–SH3 interaction has been reported.

Orientation of N- and C-terminal lobes

The major difference in the global structure of IRK and cAPK is the relative orientation of the N- and C-terminal lobes. Crystallographic studies of mammalian cAPK with and without Mg-ATP and a peptide inhibitor (PKI) have revealed two conforma-

tional states for this kinase, referred to as the open and closed forms[27] (the closed form being active). In the open form, seen in the apo and binary (+PKI) structures, the N-terminal lobe is swung away from the C-terminal lobe by 14° and translated by 0.7 Å compared with the lobes in the closed, ternary (+PKI, +Mg–ATP) structure. The N-terminal lobe of IRK is rotated 26° and translated 0.8 Å relative to closed-form cAPK and 19° and 2.1 Å relative to open-form cAPK. In addition to a rotational component that results in a wider cleft between the two lobes (axis perpendicular to the page in Fig. 2d), there is also a

---

TABLE 1   Statistics for data collection, phase determination and refinement

Data collection (20.0 to 2.5 Å)

Wavelength (Å)	Reflections* (N)	Completeness (%)	Signal $\langle I/\sigma(I)\rangle$	R_{sym}† (%)
0.9793 (remote)	22,529	95.7	22.4	2.6
1.0061 (peak)	22,461 (37,359)‡	95.4 (93.9)	21.3 (20.7)	2.5 (2.7)
1.0093 (edge)	22,346	94.9	20.9	2.5

MAD structure factor ratios§

Wavelength (Å)	Observed ratios						Scattering factors (e)	
	20.0 < d < 4.0 Å			4.0 < d < 2.5 Å				
	0.9793	1.0061	1.0093	0.9793	1.0061	1.0093	f'	f''
0.9793	0.035 (0.017)	0.019	0.034	0.037 (0.024)	0.023	0.037	−21.6	6.0
1.0061		0.054 (0.017)	0.027		0.057 (0.023)	0.032	−14.4	10.1
1.0093			0.050 (0.020)			0.057 (0.031)	−7.7	9.7

MAD phasing‖

$R(|^{o}F_{T}|) = 0.033$ $\langle \Delta(\Delta\phi)\rangle = 32.8°$ $\langle m\rangle = 0.87$
$R(|^{o}F_{A}|) = 0.315$ $\langle \sigma(\Delta\phi)\rangle = 14.6°$

Refinement¶

Model: 303 residues, 201 water molecules, 2 EMP molecules (2,588 atoms)

d-spacings (Å)	Reflections (N)	R-value (%)	Free R-value (%)	R.m.s. deviations		
				Bonds (Å)	Angles (°)	B-values (Å²)
6.0–2.1	34,747	19.6	23.2	0.011	1.6	2.2

Expression and crystallization. A recombinant baculovirus was constructed to code for human insulin receptor residues Val 978 to Lys 1,283 (L.W., S.R.H., W.A.H. and L.E., submitted). Two amino-acid substitutions were introduced: Cys 981→Ser and Tyr 984→Phe. The expressed protein was purified from lysates of baculovirus-infected insect cells (48-h post-infection) on Q-Sepharose, Superdex-200 and Mono-Q columns. Crystals of apo IRK were grown at 21 °C by vapour diffusion in hanging drops containing equal volumes of 10 mg ml^{-1} protein solution and the reservoir solution of 20% polyethylene glycol (PEG) 6000, 0.2 M malate-imidazole, pH 7.5. Macroseeding was required to grow crystals of sufficient size. The crystals belong to space group $P2_{1}2_{1}2_{1}$ and have unit cell dimensions of $a = 54.0$ Å, $b = 73.0$ Å, $c = 89.2$ Å when frozen. There is one IRK molecule in the asymmetric unit and the solvent content is 52%, assuming a partial specific volume of 0.73 cm³ g^{-1}. **Data collection.** All of the MAD data were obtained from one cryocooled mercury-derivative crystal with approximate dimensions of $0.6 \times 0.6 \times 0.07$ mm. The crystal was soaked in stabilizing solution (25% PEG 6000, 0.2 M malate–imidazole, pH 7.5) which included 0.1 mM ethylmercuric phosphate (EMP) for ~40 h, after which it was transferred to stabilizing solutions that successively included 5% and then 10% glycerol. The crystal was flash-cooled in a dry nitrogen stream at −160 °C. Data were collected on Fuji imaging plates (IPs) at three X-ray wavelengths near the mercury L_{III} edge at beamline X-4A at the National Synchrotron Light Source, Brookhaven National Laboratory. The crystal was oriented such that a* was parallel to the oscillation axis. Bijvoet pairs across the b*c* plane were collected on the same or adjacent IPs. Oscillation ranges of 1.7–1.9° and exposure times of 90–120 s were used. The exposed IPs were digitized with a Fuji scanner. **Data processing.** Raw IP data were converted to integrated intensities with DENZO[36]. ROTAVATA[37] was used to calculate scaling parameters for each IP, and these were applied with a modified version of AGROVATA[37] that does not merge redundant measurements. The MADSYS program package (W.A.H.) was used to extract phases and figures of merit. The experimental electron density map was computed using MAD-derived phases for reflections from 20.0 Å to 2.5 Å. **Model building and refinement.** Fitting of the polypeptide chain to the electron density was done using FRODO[38] on a Silicon Graphics Iris workstation. Reference was made to the cAPK structure[19] during model building. COMBIN (W.A.H.) was used to combine the experimental phases with phases calculated from the partial model. Least-squares and simulated annealing refinement were done using X-PLOR[39], gradually extending the resolution from 2.5–2.1 Å. When the model was nearly complete, simulated annealing omit maps were computed in which ten consecutive residues were omitted from the calculation. As defined in PROCHECK[40], there are no residues in disallowed main-chain torsion angle regions and two residues, Arg 1,131 and Asp 1,156, in generously allowed regions. Side chains for the following residues were not modelled past Cβ owing to poor electron density: 981, 987, 1,034, 1,047, 1,096, 1,127, 1,153–1,157, 1,159 and 1,237. The average B-value for protein atoms is 23.2 Å². The side chains for Met 1,051 and Met 1,076 have been modelled in two different conformations. Water molecules whose B-values refined to >50 Å² were omitted from the subsequent round of refinement. EMP is bound to Cys 1,056 and Cys 1,234.
  * Unique reflections for acentrics in point group symmetry P222 and centrics in Pmmm.
  † $R_{sym} = 100 \times \Sigma_{hkl} \Sigma_i |I_i - \langle I\rangle|/\Sigma_{hkl} \Sigma_i I_i$, where I_i is the ith measurement and $\langle I\rangle$ is the weighted mean of all measurements of I.
  ‡ Values in parentheses are for data from 20.0–2.1 Å used for refinement.
  § Table values represent $\langle\Delta|F|^2\rangle^{1/2}/\langle|F|^2\rangle^{1/2}$, where $\Delta|F|$ is the absolute value of the Bijvoet difference at one wavelength (diagonal elements) or of the dispersive difference at two wavelengths (off-diagonal elements). The values in parentheses are the ratios for centric Bijvoet reflections, which would be equal to zero for perfect data and serve as an estimate of the noise in the anomalous signals. The listed scattering factors are the refined values for mercury.
  ‖ $R = \Sigma_{hkl} \Sigma_i |F_i - \langle|F|\rangle|/\Sigma_{hkl}|F|$. $^{o}F_{T}$ is the structure factor due to normal scattering from all the atoms. $^{o}F_{A}$ is the structure factor due to normal scattering from the anomalous scatterers only, and $\Delta\phi$ is the phase difference between $^{o}F_{T}$ and $^{o}F_{A}$. $\Delta(\Delta\phi)$ is the difference between two independent determinations of $\Delta\phi$. Values given are based on calculations that did not include reflections with $m = 0$; $\langle m\rangle$ is the mean figure of merit including reflections with $m = 0$.
  ¶ A subset of the data (10%) was excluded from the refinement and used for the free R-value calculation until the final round of refinement, in which all of the data ($F > 2\sigma$) were used. R-value $= 100 \times \Sigma_{hkl}||F_o| - |F_c||/\Sigma_{hkl}|F_o|$. R.m.s. deviation in B-values is for bonded atoms.

rotational component along the long axis of the molecule (vertical in Fig. 2d, anticlockwise from above upon opening). The apo forms of MAPK and twitchin kinase also adopt open conformations with lobe rotations of 17° and 30° (refs 21 and 22, respectively), vis-à-vis closed-form cAPK, whereas apo CDK2 adopts a closed conformation[20].

The N- and C-terminal lobes of IRK are held apart by steric interactions between conserved Gly 1,005 of the glycine-rich, nucleotide-binding loop (between β1 and β2) and residues Phe 1,151 and Gly 1,152 of the kinase-conserved Asp-Phe-Gly sequence, near the beginning of the activation loop (Figs 2b, 5b). In contrast, the interaction of residues in αC with Asp-
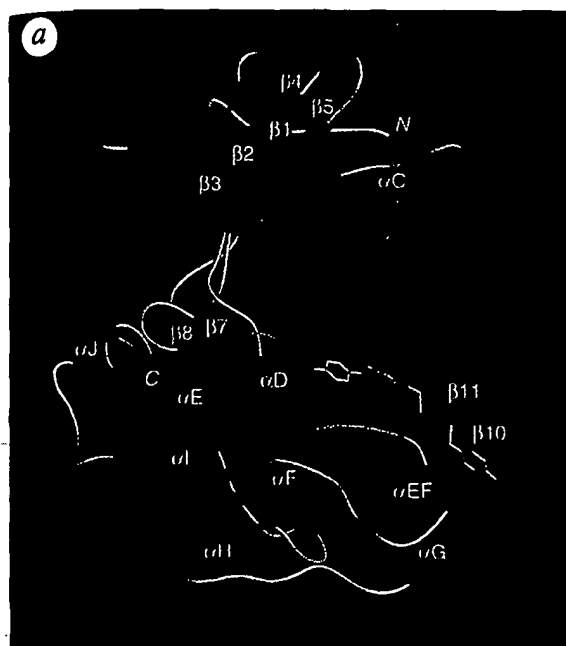
FIG. 2 a, Ribbon diagram of the apo IRK structure. The α-helices are shown in red, the β-strands in green, the side chains of tyrosines 1,158, 1,162 and 1,163 in yellow, the glycine-rich, nucleotide-binding loop in orange, the catalytic loop in dark blue, the activation loop in violet, the P+1 loop in light blue, and the kinase insert region in dark grey. The termini are denoted by N and C. b, Stereo view of a Cα trace of IRK in the same orientation as in a. Every tenth residue is marked with a filled circle and every twentieth residue is labelled. c, Cα trace of the N-terminal lobes of IRK (residues 993–1,081) and cAPK (residues 40–125) in which the Cα atoms of the β-sheet cores are superposed. IRK is in red, cAPK in blue. The view is rotated slightly from that in a. d, Stereo view of a Cα trace of IRK (residues 981–1,283) and cAPK (residues 40–310) in which the Cα atoms of αD–αH are superposed; colouring as in c. The activation loop in IRK (residues 1,149–1,170) and the corresponding residues in cAPK (183–199) are shown in orange and green, respectively. The view is the same as in b. Brookhaven Protein Data Bank entry 1ATP for cAPK[42] was used in the Cα superpositions. Preparation of figures: a made with RIBBONS[43]; b with MOLSCRIPT[44]; c and d with GRASP[45].
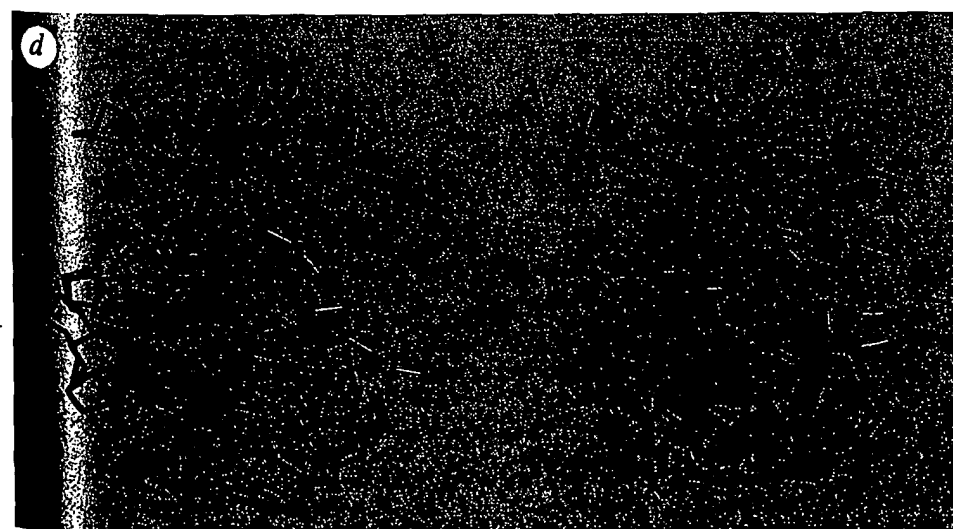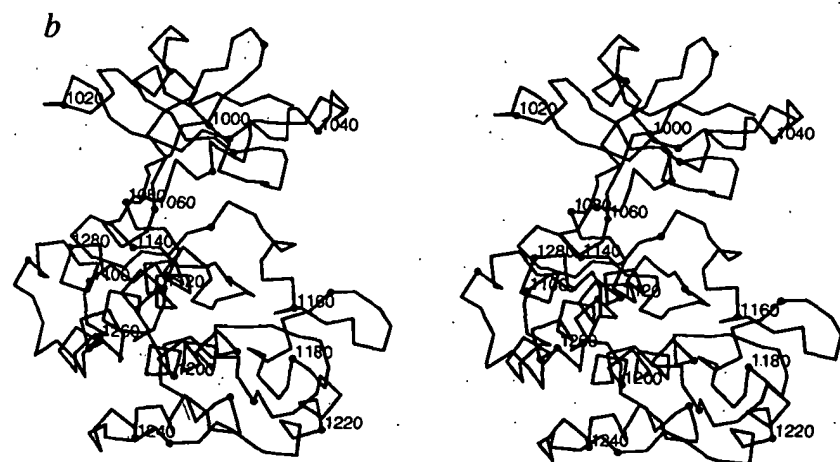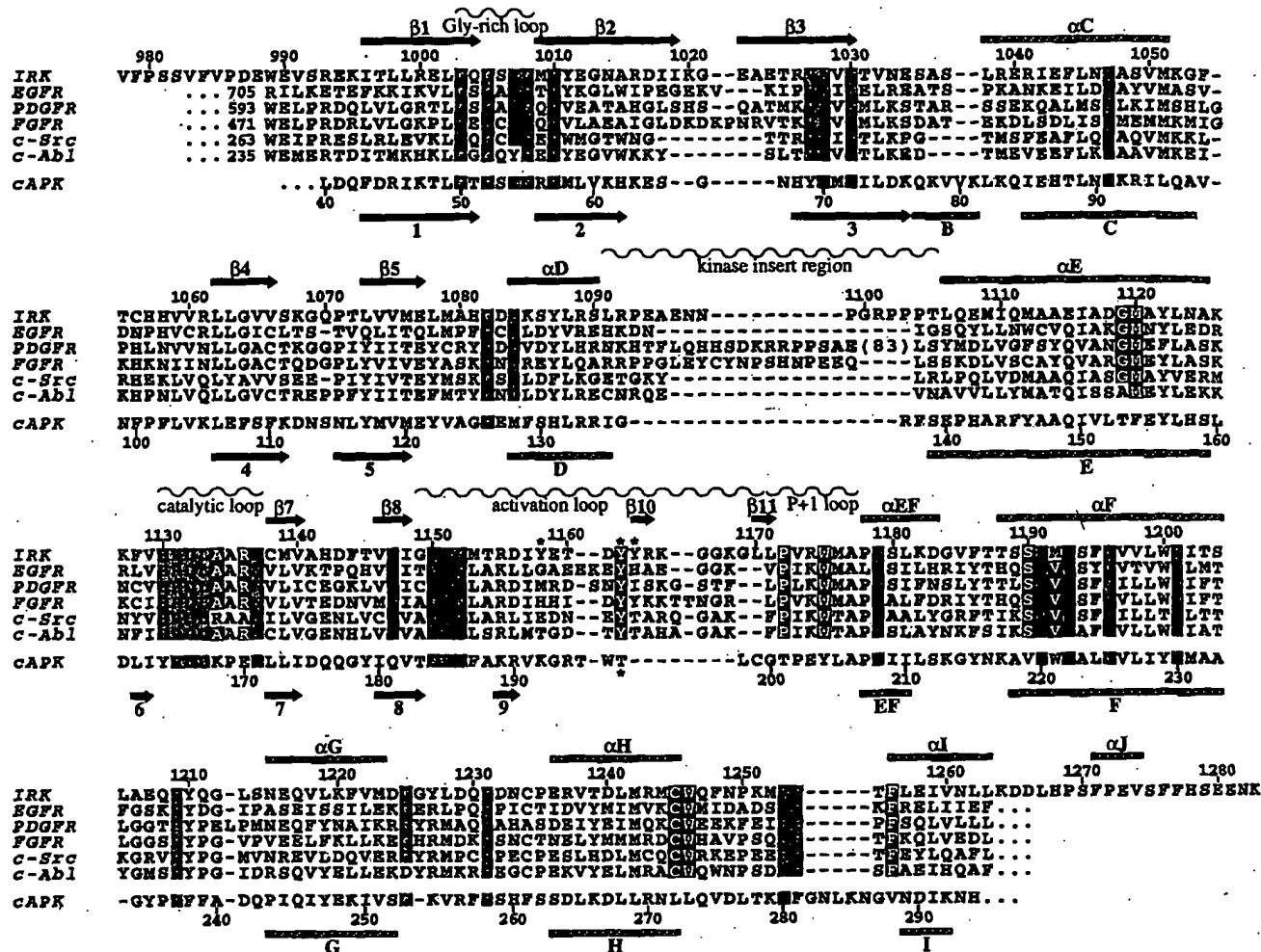
FIG. 3 Structure-based sequence alignment of IRK, cAPK, and the tyrosine kinase domains of the EGF, PDGF (β) and FGF (FLG) receptors and of c-Src and c-Abl. The secondary structure assignments for IRK and cAPK were obtained using the Kabsch and Sander algorithm[46] as implemented in PROCHECK[40]. A residue is highlighted if found at that position in >90% of the tyrosine kinase sequences listed in Hanks[47]. Of these, the residues that show conservation in the PSK family as well are shaded, whereas those that are PTK-specific are shown in reverse contrast. The tyrosine autophosphorylation sites in IRK and the essential phosphorylation site in cAPK are marked with an asterisk. No attempt was made to align the PTK sequences between αD and αE and between β2 and β3.

Phe-Gly accounts for the open conformation of MAPK[21]. The disposition of the lobes in apo IRK places kinase-conserved Lys 1,030, a residue in the N-terminal lobe implicated in ATP binding, a comparatively large distance from the catalytic residues in the C-terminal lobe. In IRK the distance from Lys 1,030 to Asp 1,132, the catalytic base, is 13.7 Å (Nζ–Oδ2), compared with 12.4 Å in MAPK[21], 11.2 Å in open-form cAPK, and 7.8 Å in closed-form cAPK.

Active site

The roles of the highly conserved residues in the protein kinase family have been largely elucidated from the crystallographic studies of cAPK complexed with Mg-ATP and PKI[28,29] (reviewed in refs 30 and 31). The so-called catalytic loop lies between β-strands 6 and 7 in cAPK. In this loop, Asp 166 (Asp 1,132) and Asn 171 (Asn 1,137) are nearly invariant in both the PSK and PTK families, and Lys 168 (Ala 1,134) is highly conserved in the PSK family. Asp 166 is the catalytic base in the phosphotransfer reaction; Lys 168 provides charge neutralization, and Asn 171 is hydrogen-bonded to Asp 166O and is involved in Mg^{2+} coordination. In the PTK family, an arginine is present in the catalytic loop rather than a lysine, either two (Src subfamily) or four residues (all other PTKs; Arg 1,136 in IRK) from the catalytic base (Fig. 3).

One of the most striking features of the apo IRK structure is the presence of Tyr 1,162, a tyrosine that is autophosphorylated upon insulin binding, in the active site of the enzyme, seemingly poised for cis-autophosphorylation (Fig. 4a, b). The hydroxyl group of the Tyr 1,162 phenolic ring is hydrogen-bonded to the carboxylate group of the catalytic base, Asp 1,132 (Oη–Oδ2: 2.6 Å), and to the guanidinium group of Arg 1,136 (Oη–Nε: 2.9 Å). The phenolic ring is oriented in part by the pyrrolidine ring of PTK-conserved Pro 1,172. The distance between the centroid of the pyrrolidine ring and the edge of the phenolic ring is 3.6 Å. Arginine 1,136 makes several other contacts: salt bridges to the carboxylate groups of Asp 1,132 and Asp 1,161, and an axial polar interaction[32] with the indole ring of Trp 1,175 (Nη1–centroid of six-membered ring: 3.1 Å).

The catalytic loop conformations of IRK and cAPK are very similar despite several sequence differences (Fig. 4b). A superposition of the catalytic loops reveals that Arg 1,136Nε is the counterpart to Lys 168Nζ in cAPK. The arginine in the catalytic loop of PTKs appears to serve a dual purpose: it provides charge neutralization at the phosphotransfer site (like Lys 168 in cAPK), and makes a hydrogen bond (via Nη2) with Oδ1 of the catalytic base, which in cAPK is hydrogen-bonded to Thr 201Oγ1. Either a threonine or serine is found in the PSK
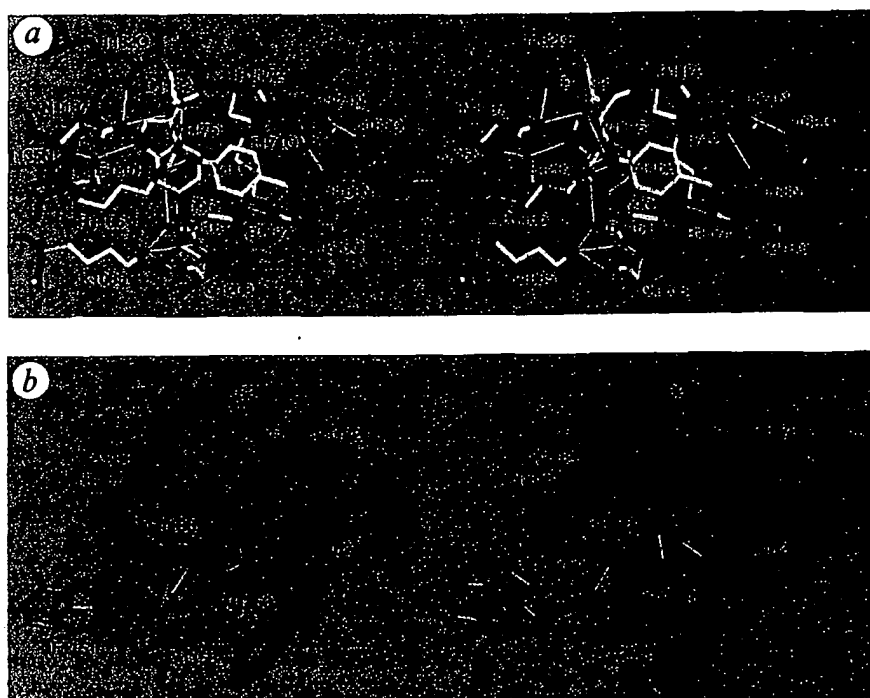
FIG. 4 a, Stereo view of the active site of IRK. Highlighted side-chain and main-chain atoms are colour-coded by atom type: carbon, yellow; nitrogen, light blue; oxygen, red. All other atoms are shown in dark green. Hydrogen-bonding interactions are shown by white lines. The view for a and b is approximately perpendicular to that in Fig. 2a, looking along the long axis of the molecule from the N- to the C-terminal lobe. b, Stereo view of the active sites of apo IRK and ternary cAPK in which the catalytic loops are superposed. IRK is in red, cAPK in blue, ATP and PKI from ternary cAPK in orange and green, respectively. Alanine 21 of PKI was changed to serine with the side-chain dihedral angle set to −60°. The labelled IRK residues and the corresponding cAPK residues (in parentheses) are Asp 1,132 (Asp 166), Ala 1,134 (Lys 168), Arg 1,136 (Glu 170), Asn 1,137 (Asn 171), Phe 1,151 (Phe 185), Met 1,153, Tyr 1,158, Tyr 1,162, Pro 1,172 (Thr 201) and Trp 1,175 (Tyr 204). Because of side-chain disorder, only Cβ of Met 1,153 is shown. Figures drawn with GRASP[45].
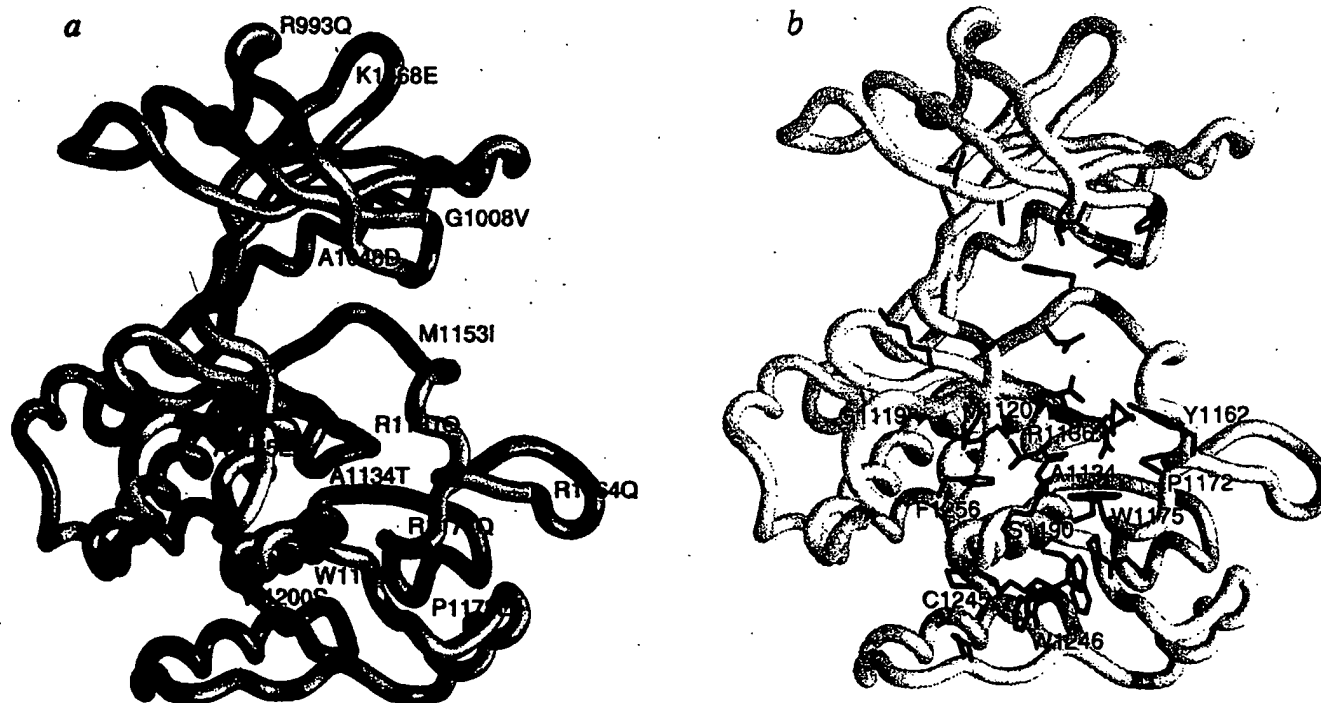
FIG. 5 a, Missense mutations in NIDDM patients mapped onto the IRK structure. The mutations, whose positions are shown in red, are Arg 993→Gln, Gly 1,008→Val, Ala 1,048→Asp, Lys 1,068→Glu, Arg 1,131→Gln, Ala 1,134→Thr, Ala 1,135→Glu, Met 1,153→Ile, Arg 1,164→Gln, Arg 1,174→Gln, Pro 1,178→Leu, Trp 1,193→Leu and Trp 1,200→Ser. b, Mapping of the highly conserved residues in the PTK family. In green are residues that show conservation in the PSK family as well, and in red are residues that are PTK-specific (as in Fig. 3). PTK-specific residues are labelled. The positions of conserved glycines are indicated by colouring (green or red) of the backbone representation. Figures drawn with GRASP[45].

family at cAPK position 201, whereas a proline (Pro 1,172) is found in the PTK family.

Although the hydroxyl group of Tyr 1,162 is in position for phosphotransfer with respect to the catalytic loop residues, the ATP-binding site appears to be inaccessible, precluding *cis*-autophosphorylation of Tyr 1,162. When the catalytic loops of apo IRK and ternary cAPK are superposed, Gly 1,152 and Met 1,153 of the IRK activation loop intersect with (superposed) ATP near the α- and β-phosphates (Fig. 4b). Furthermore, the side chain of Phe 1,151 occupies the hydrophobic pocket in which the adenine ring of ATP sits in ternary cAPK. Val 57 and Leu 173 are situated on either side of the adenine ring in ternary cAPK, and in apo IRK the corresponding residues, Val 1,010 and Met 1,139 (most often a leucine in the PTK family), flank the phenolic ring of Phe 1,151. In both the open and closed forms of cAPK, the phenolic ring of Phe 185 (Phe 1,151) resides in a hydrophobic environment formed by Leu 95 (Met 1,051) and Tyr 164 (His 1,130). The corresponding pocket in the apo IRK structure is largely unoccupied, but is presumably filled by Phe 1,151 in the phosphorylated, activated form of IRK.

Tyrosine substrate selectivity

The polypeptide chain in the near vicinity of Tyr 1,162 appears to mimic the way in which a peptide substrate binds; Tyr 1,162 is an autophosphorylation site (P-site), albeit in *trans* (discussed below), and the interactions of Asp 1,161 (P-1 residue) and Tyr 1,163 (P+1 residue) are consistent with observed substrate specificities for IRK (see below). Therefore, the determinants of tyrosine versus serine/threonine substrate selectivity can be addressed from this apo structure. As seen in the superposition in Fig. 4b, the hydroxyl groups of Tyr 1,162 and Ser 21 of PKI (serine substituted for alanine at the P-site) occupy nearly the same position. Clearly, the hydroxyl group of a serine or threonine side chain at this position in IRK (and in a peptide substrate) would be too short to reach the phosphotransfer site.

The position of the main chain at Tyr 1,162 is determined by the loop that contains residues Leu 1,171 to Ala 1,177, referred to as the P+1 loop (in cAPK, the loop with which the P+1 residue of PKI interacts[31]). Alignments of protein kinase sequences have revealed that the sequence corresponding to Pro 1,172 to Trp 1,175 in IRK is characteristic of the PTK family[33]. The two main-chain interactions that govern the distance from the Cα atom of Tyr 1,162 to the phosphotransfer site are hydrogen bonds between Tyr 1,163N and Leu 1,171O and between Tyr 1,163O and Leu 1,171N (Fig. 4a). Similar main-chain interactions are present between the P+1 residue of PKI and Gly 200 (Leu 1,171) in ternary cAPK. The active site superposition in Fig. 4b shows that the conformation of the P+1 loop, especially near Pro 1,172 (Thr 201), is an important determinant in substrate selectivity.

The conformation of the IRK P+1 loop is stabilized by a number of interactions. Leu 1,171, Val 1,173 and Met 1,176 form a hydrophobic cluster together with Leu 1,181 and Leu 1,219. Arg 1,174Nη1 and Nη2 are hydrogen-bonded to Pro 1,209O and Leu 1,213O in the connection between αF and αG. An arginine or a lysine is found at position 1,174 in all but the Tyk/Jak subfamily of PTKs. Conserved Trp 1,175 plays a fundamental role in coupling the P+1 loop to the catalytic loop. In addition to the direct, axial polar interaction with Arg 1,136, there is an indirect interaction mediated by Glu 1,201; Glu 1,201Oε2 is hydrogen-bonded to both Trp 1,175Nε1 and Ala 1,135N (Fig. 4a). Glu 1,201 is highly conserved in the PTK family, less so in the PSK family. The primary coupling between the P+1 and catalytic loops of cAPK is the interaction of Thr 201 with Asp 166 and Lys 168, which positions the P+1 loop nearer to the catalytic loop than in IRK, contributing to serine/threonine versus tyrosine substrate selection.

Peptide substrate specificity

The interactions of Asp 1,161 (P-1 residue) and Tyr 1,163 (P+1 residue) with other IRK residues correlates well with observed IRK substrate preferences. Asp 1,161 is salt-bridged to Lys 1,085 and Arg 1,136 and hydrogen-bonded to Gln 1,208 (Fig. 4a). Lys 1,085 and two other positively charged residues, Arg 1,089 and Lys 1,092, all lie along the same face of αD. The latter two residues are potential contacts for negatively charged P-2 and P-3 residues of a peptide substrate. The side chain of Tyr 1,163 (also an autophosphorylation site) resides in a hydrophobic environment formed by Val 1,173 and Leu 1,219, and the hydroxyl group is hydrogen-bonded to the carboxylate group of Glu 1,216. IRS-1 is phosphorylated on at least eight tyrosines by the insulin receptor[14]. Four of these tyrosines are preceded by an aspartate or glutamate and all eight are followed by a hydrophobic residue. Furthermore, in a study of PTK substrate specificities using an *in vitro* combinatorial approach, the preferred peptide substrate for IRK contained the sequence Glu-Glu-Glu-Tyr-Met-Met-Met (Z. Songyang and L. Cantley, unpublished results).

IRK mutations in NIDDM

The missense mutations in the tyrosine kinase portion of the insulin receptor gene that have been found in patients with NIDDM[17] are mapped onto the IRK structure in Fig. 5a. Many of these mutations involve replacement of hydrophobic residues with others of different size or with charged residues. At the highly conserved Gly 1,008 position, no room for a valine side-chain exists owing to the proximity of the main chain of residues 1,031–1,033. In αC, Ala 1,048Cβ sits in a hydrophobic pocket which could not accommodate the larger aspartate side chain. Similarly, Ala 1,134→Thr and Ala 1,135→Glu in the catalytic loop and Pro 1,178→Leu at the start of αEF would result in steric clashes. The two tryptophan mutations, Trp 1,193→Leu and Trp 1,200→Ser, apparently destabilize the hydrophobic packing in the C-terminal lobe by introducing side chains of smaller size. These two tryptophans, along with PTK-conserved tryptophans 1,175 and 1,246, form a cluster of four tryptophans in this region of the C-terminal lobe.

Four Arg→Gln mutations have been identified. Both Nη1 and Nη2 of the Arg 993 guanidinium group are hydrogen-bonded to the carbonyl oxygen of Pro 1,071 in the loop between β4 and β5, thereby stabilizing that region of the N-terminal lobe. Pro 1,071 is a *cis* proline, the only one in the structure, and is reasonably well conserved in the PTK family. In the PTK sequences examined, there is a strong correlation between proline at position 1,071 and arginine at position 993 (Fig. 3). The guanidinium group of Arg 1,131 makes no electrostatic contacts in the present structure, yet this residue is invariant in the PTK family. Based on the cAPK structure, it is expected that in the phosphorylated form of IRK, Arg 1,131 will be salt-bridged to one of the phosphotyrosines in the activation loop (see below). As already discussed, the guanidinium group of Arg 1,174 in the P+1 loop is hydrogen-bonded to Pro 1,209O and Leu 1,213O. Loss of the latter interaction upon substitution with a glutamine probably destabilizes the conformation of the P+1 loop, which is important in substrate positioning. The present structure does not provide an explanation for the adverse effects of Arg 1,164→Gln, nor of Met 1,153→Ile and Lys 1,068→Glu. Insights into the two activation loop mutations are expected from a structure of the phosphorylated form of IRK.

Other conserved residues

A number of highly conserved residues in the PTK family (Figs 3, 5b) do not play a direct role in catalysis, but are important instead in structure stabilization. Many of these residues are also conserved in the PSK family. As in cAPK, Glu 1,179 (Glu 208) is salt-bridged to Arg 1,253 (Arg 280), and Asp 1,191 (Asp 220) is hydrogen-bonded to the backbone amide groups of His 1,130 (Tyr 164) and Arg 1,131 (Arg 165) of the catalytic loop. The

Glu 1,047 (Glu 91) side chain is not salt-bridged to Lys 1,030 (Lys 72), but rather is disordered. Glycines 1,082, 1,152 and 1,225 have backbone torsion angles less favourable for $C\beta$-containing residues. Gly 1,196 is in an α-helical conformation, but a $C\beta$ atom at this position would result in steric clashes with the side chains of Cys 1,245 and Trp 1,246. The imidazole group of His 1,130 is involved in two electrostatic interactions: $N\delta1$ is hydrogen-bonded to the backbone amide group of Asp 1,132 (Asp 166), and $N\epsilon2$ is hydrogen-bonded to a well ordered water molecule which in turn is hydrogen-bonded to Asn 1,137O (Asn 171) and Asp 1,150O$\delta2$ (Asp 184).

Conserved, PTK-specific residues that play a role in structure stabilization (and not mentioned previously) include Gly 1,119, Met 1,120, Ser 1,190, Cys 1,245 and Phe 1,256. A $C\beta$ atom at Gly 1,119 in αE would result in a steric clash with the carbonyl oxygen of His 1,058. Met 1,120 and Phe 1,256 form part of the hydrophobic environment in which Leu 1,133 of the catalytic loop is situated. The hydroxyl group of Ser 1,190 is hydrogen-bonded to both the amide nitrogen and carbonyl oxygen of Thr 1,187, as well as to a water molecule which is also hydrogen-bonded to Asp 1,191O$\delta2$. The environment of the Cys 1,245 side chain in αH is hydrophobic, yet it is hydrogen-bonded to the carbonyl oxygen of Leu 1,241, also in αH (Sγ-O: 3.4 Å). The same hydrogen-bonding interaction for a buried cysteine in an α-helix was observed in myohaemerythrin[34].

Cis-inhibition and trans-activation

The apo IRK structure reveals a novel mechanism of autoinhibition in which the hydroxyl group of Tyr 1,162 is bound in the active site. Several PSKs, including protein kinase C, calmodulin-dependent protein kinase II and myosin light-chain kinase, contain within the same polypeptide chain a pseudosubstrate sequence that blocks access to the active site[35]. Pseudosubstrate sequences resemble exogenous substrate sequences, with the P-site serine or threonine most often replaced by alanine. In contrast, IRK possesses not a pseudosubstrate but a bona fide substrate known to be autophosphorylated in response to insulin.

Although evidence for cis-autophosphorylation has been reported[9,10], our autophosphorylation experiments on IRK (L.W., S.R.H., W.A.H. and L.E., manuscript submitted) and those of others on the insulin receptor cytoplasmic domain[11] and intact receptor[12] are consistent with a trans-autophosphorylation reaction, including the first phosphorylation event. Moreover, studies on the receptors for EGF, PDGF and fibroblast growth factor (FGF) indicate that autophosphorylation occurs via a trans mechanism[4]. Based on these results and the observation that the ATP-binding site is blocked in the unphosphorylated IRK structure, we conclude that the binding of ATP and of 'self' Tyr 1,162 in the active site are mutually exclusive, such that cis-autophosphorylation of Tyr 1,162 does not occur to any appreciable extent. Further evidence for this binding exclusivity comes from limited tryptic digestion experiments on IRK which demonstrate that the activation loop is more readily proteolysed in the presence of a non-hydrolysable ATP analogue (data not shown).

A model for insulin receptor activation emerges from the apo IRK and ternary cAPK structures and the biochemical data. In solution an equilibrium exists between two activation loop conformations, one in which Tyr 1,162 is engaged in the active site and both substrate and ATP-binding sites are inaccessible, as seen in the present structure, and the other in which Tyr 1,162 is disengaged and both binding sites are accessible. The former conformation is highly favoured in the unphosphorylated state. When Tyr 1,162 is disengaged, Mg-ATP can bind if it is present. The kinase is then transiently active until Tyr 1,162 returns to the active site. In the absence of insulin, the disposition of the receptor's cytoplasmic domains is such that trans-autophosphorylation is prevented. When insulin binds to the α-chains, a change within the quaternary structure of the receptor places the phosphorylation sites of one β-chain within reach of the

active site of the other β-chain, the juxtamembrane tether providing the required flexibility. Intramolecular, trans-autophosphorylation can then occur when Tyr 1,162 is disengaged and Mg-ATP is bound. If phosphorylation occurs on an activation loop tyrosine, the loop equilibrium is shifted towards a non-inhibiting conformation—Tyr 1,162 disengaged from the active site—which is stabilized by specific electrostatic interactions involving the phosphotyrosine. The result is a significant increase in kinase activity (up to 200-fold for the tri-phosphorylated state in vitro; L.W., S.R.H., W.A.H. and L.E., manuscript submitted).

The activation loop of virtually all PTKs contains from one to three tyrosines, one of which can be readily aligned in sequence with Tyr 1,162 (Fig. 3). Also, Arg 1,131 is invariant in the PTK family and the equivalent arginine in cAPK, Arg 165, is salt-bridged to phosphorylated Thr 197 (P-Thr 197) in the cAPK activation loop. We propose that phosphorylation of Tyr 1,162 is the key step in insulin receptor kinase activation, whereby P-Tyr 1,162 will be salt-bridged to Arg 1,131, stabilizing the non-inhibiting conformation of the activation loop. P-Tyr 1,162 may also interact with Arg 1,155, as this residue is either an arginine or lysine in nearly all PTKs and the corresponding lysine in cAPK (Lys 189) interacts with P-Thr 197. The functions of P-Tyr 1,158 and P-Tyr 1,163 in activation are less clear. A crystal structure of the phosphorylated form of IRK should provide insights into the role of phosphorylation in kinase activation.

We believe that the main aspect of this cis-inhibition/trans-activation mechanism will apply to many PTK family members. In the unphosphorylated state, the activation loop tyrosine corresponding to Tyr 1,162 will be bound in the active site and, concomitantly, ATP binding will be blocked. Autophosphorylation of this tyrosine in trans will stabilize the non-inhibiting conformation of the activation loop through electrostatic interactions between the phosphotyrosine and positively charged residues. Activation of the insulin receptor and its subfamily members will differ from that of other receptor PTKs in the mechanism of signal transduction—ligand-induced intramolecular rearrangement rather than ligand-induced oligomerization. □

1. White, M. F. & Kahn, C. R. J. biol. Chem. 269, 1–4 (1994).
2. Ullrich, A. et al. Nature 313, 756–761 (1985).
3. Ebina, Y. et al. Cell 40, 747–758 (1985).
4. Ullrich, A. & Schlessinger, J. Cell 61, 203–212 (1990).
5. Tornqvist, H. E. et al. J. biol. Chem. 263, 350–359 (1988).
6. Tavare, J. M., O'Brien, R. M., Siddle, K. & Denton, R. M. Biochem. J. 253, 783–788 (1988).
7. White, M. F., Shoelson, S. E., Keutmann, H. & Kahn, C. R. J. biol. Chem. 263, 2969–2980 (1988).
8. Kohanski, R. A. Biochemistry 32, 5773–5780 (1993).
9. Villalba, M. et al. Proc. natn. Acad. Sci. U.S.A. 86, 7848–7852 (1989).
10. Shoelson, S. E., Boni-Schnetzler, M., Pilch, P. F. & Kahn, C. R. Biochemistry 30, 7740–7746 (1991).
11. Cobb, M. H., Sang, B.-C., Gonzalez, R., Goldsmith, E. & Ellis, L. J. biol. Chem. 264, 18701–18706 (1989).
12. Frattali, A. L., Treadway, J. L. & Pessin, J. E. J. biol. Chem. 267, 19521–19528 (1992).
13. Koch, C. A., Anderson, D., Moran, M. F., Ellis, C. & Pawson, T. Science 252, 668–674 (1991).
14. White, M. F. Curr. Opin. gen. Dev. 4, 47–54 (1994).
15. Ebina, Y. et al. Proc. natn. Acad. Sci. U.S.A. 84, 704–708 (1987).
16. Chou, C. K. et al. J. biol. Chem. 262, 1842–1847 (1987).
17. Accili, D. et al. J. Endocr. Invest. 15, 857–864 (1992).
18. Hendrickson, W. A. Science 254, 51–58 (1991).
19. Knighton, D. R. et al. Science 253, 407–414 (1991).
20. DeBondt, H. L. et al. Nature 363, 595–602 (1993).
21. Zhang, F., Strand, A., Robbins, D., Cobb, M. H. & Goldsmith, E. J. Nature 367, 704–711 (1994).
22. Hu, S.-H. et al. Nature 369, 581–584 (1994).
23. Vernon, M., Radzio-Andzelm, E., Tsigelny, I., Ten Eyck, L. F. & Taylor, S. S. Proc. natn. Acad. Sci. U.S.A. 90, 10618–10622 (1993).
24. Ren, R., Mayer, B. J., Cicchetti, P. & Baltimore, D. Science 259, 1157–1161 (1993).
25. Musacchio, A., Saraste, M. & Wilmanns, M. Nature struct. Biol. 1, 546–551 (1994).
26. Yu, H. et al. Cell 76, 933–945 (1994).
27. Zheng, J. et al. Protein Science 2, 1559–1573 (1993).
28. Zheng, J. et al. Biochemistry 32, 2154–2161 (1993).
29. Bossemeyer, D., Engh, R. A., Kinzel, V., Ponstingl, H. & Huber, R. EMBO J. 12, 849–859 (1993).
30. Wei, L., Hubbard, S. R., Smith, R. F. & Ellis, L. Curr. Opin. struct. Biol. 4, 450–455 (1994).
31. Taylor, S. T. & Radzio-Andzelm, E. Structure 2, 345–355 (1994).
32. Burley, S. K. & Petsko, G. A. FEBS Lett. 203, 139–143 (1986).
33. Hanks, S. K., Quinn, A. M. & Hunter, T. Science 241, 42–52 (1988).

34. Sheriff, S., Hendrickson, W. A. & Smith, J. L. J. molec. Biol. 197, 273–296 (1987).
35. Kemp, B. E. & Pearson, R. B. Biochim. biophys. Acta 1094, 67–76 (1991).
36. Otwinowski, Z. in Data Collection and Processing (eds Sawyer, L., Isaacs, N. & Bailey, S.) 56–62 (SERC Daresbury Laboratory, Warrington, UK, 1993).
37. CCP4: A Suite of Programs for Protein Crystallography (SERC Collaborative Computing Project no. 4, Daresbury Laboratory, Warrington, UK, 1979).
38. Jones, T. A. Meth. Enzym. 115, 157–171 (1985).
39. Brünger, A. T. X-PLOR, Version 3.1, A System for X-ray Crystallography and NMR (Yale Univ. Press, New Haven, USA, 1992).
40. Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. J. appl. Crystallogr. 26, 283–291 (1993).
41. Evans, S. V. J. molec. Graph. 11, 134–138 (1993).
42. Zheng, J. et al. Acta crystallogr. D49, 362–365 (1993).
43. Carson, M. J. molec. Graph. 5, 103–106 (1987).
44. Kraulis, P. J. J. appl. Crystallogr. 24, 946–950 (1991).
45. Nicholls, A., Sharp, K. A. & Honig, B. Proteins Struct. Funct. Genet. 11, 281–296 (1991).
46. Kabsch, W. & Sander, C. Biopolymers 22, 2577–2637 (1983).
47. Hanks, S. K. Curr. Opin. struct. Biol. 1, 369–383 (1991).

ACKNOWLEDGEMENTS. We thank H. Yamaguchi, M. Cuff, L. Shapiro and C. Ogata for help in synchrotron data collection; N. McDonald and H. Yamaguchi for discussion; J. Schlessinger for critical reading of the manuscript; N. Belgado for tissue culture; A. Nicholls for help in the use of GRASP; and L. Ten Eyck for making the coordinates of mammalian, binary cAPK available. This work was supported in part by grants from the NIH (W.A.H. and L.E.), the NSF (W.A.H.) and the W. M. Keck Foundation (L.E.). L.W. is supported by a Juvenile Diabetes Foundation International postdoctoral fellowship. Beamline X-4A at the National Synchrotron Light Source, a DOE facility, is supported by the Howard Hughes Medical Institute. Coordinates have been deposited in the Brookhaven Protein Data Bank.

Evidence for magnetic-field-induced anisotropy of the interstellar medium

K. M. Desai*†, C. R. Gwinn* & P. J. Diamond†

* Physics Department, University of California, Santa Barbara, California 93106, USA
† National Radio Astronomy Observatory, PO Box 0, Socorro, New Mexico 87801, USA

TURBULENCE in the interstellar medium transfers energy from parsec-sized regions to much smaller scales, and may be responsible for supporting clouds against gravitational collapse[1]. Fluctuations in the electron density, which trace turbulence, occur on scales ranging from 10^6 to $>10^{13}$ cm—the largest range of spatial scales seen in natural turbulence. Despite almost thirty years of study, however, the causes and effects of interstellar turbulence are still poorly understood. Here we present observations of OH masers in the Galactic star-forming complex W49N, which we use as point sources to investigate scattering along the line of sight. The masers' images are elliptical, and aligned roughly perpendicular to the Galactic plane. This alignment suggests that the magnetic field of our Galaxy influences interstellar turbulence[2,3] by mediating the transfer of energy from large to small spatial scales.

We chose to observe the OH masers in W49N because this 11-kpc line of sight samples interstellar turbulence over a long path through the plane of the Galaxy, and so yields its average properties. Moreover, the expected angular broadening due to scattering of ~45 milliarcseconds (mas), as extrapolated from measurements of H_2O masers in W49N (ref. 4) and from OH masers on other lines of sight towards other directions[5,6], is fairly well matched to the excellent imaging capabilities of the recently completed Very-Long-Baseline Array (VLBA). Because masers are extremely bright, pointlike objects, their scatter-broadened images, known as scattering disks, are unaffected by their intrinsic structure, or sensitivity limitations[4]. Often many masers lie close together in star-forming regions, so that many close but distinct lines of sight can be studied simultaneously.

We observed the emission from the W49N OH masers at 1,667 MHz on 13 March 1993, using very-long-baseline interferometry at eight VLBA antennas and one antenna of the Very Large Array (VLA). We processed the data with the Mark II processor[7] of the National Radio Astronomy Observatory to obtain cross-power spectra for each antenna pair, with spectral resolution corresponding to 0.156 km s^{-1} in Doppler velocity. We analysed these data with the AIPS software package. We used observations of unresolved continuum sources and bright maser features to remove instrumental and atmospheric gain and phase variations at each antenna. We analysed, and discuss in this Letter, only frequency channels in which confusing emis-

sion from OH masers in the neighbouring source W49S was absent, and in which masers were detected at five or more times the noise level. We fit elliptical models for scattering disks directly to the measured cross-power spectra, and determined the number and location of maser spots empirically.

The scattering disks of the masers are elongated, and are approximately aligned perpendicular to the Galactic plane. Figure 1 shows the observed distribution of maser spots on the sky. The contours for each maser spot at half of the maximum brightness are shown, magnified eight times for clarity. The mean position angle of the minor axes is parallel to the Galactic plane to within 5°. The alignment of the spots is highly significant, with some evidence for systematic variation across the source. Figure 2 shows the fitted major- and minor-axis sizes, with error bars from post-fit residuals. The axial ratios range from 1.5:1
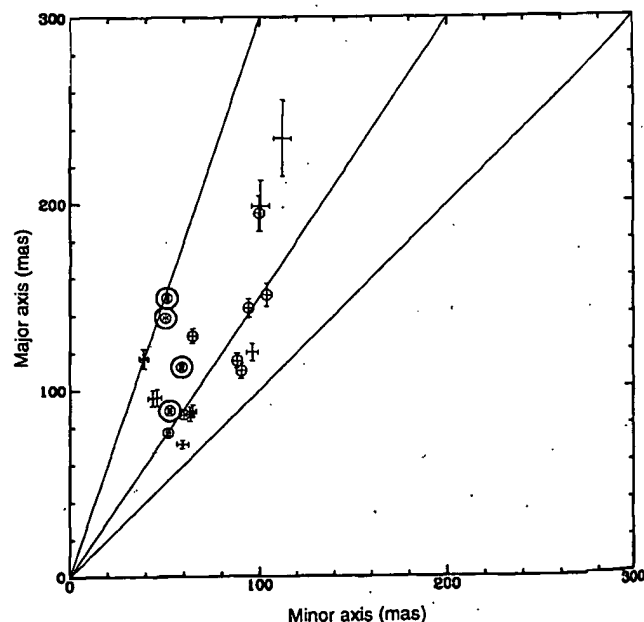

FIG. 1 The observed distribution of maser spots on the sky, showing ~2:1 elongation, and alignment of minor axes with the Galactic plane, for all maser spots detected at ≥5 times the noise level. The average Galactic magnetic field runs along lines of constant Galactic latitude, from lower right to upper left. Contours are 50% of peak brightness (magnified eight times for clarity). We note a systematic variation of the position angles of the major axes with increasing Galactic latitude. The crossed ellipse at the lower right of the figure indicates the (also eight times magnified) effective resolution element, oriented 40° away from the Galactic plane with an axial ratio of 1.4:1, obtained from our many interferometer baselines. (Dec., declination, RA, right ascension; b, Galactic latitude; l, Galactic longitude).
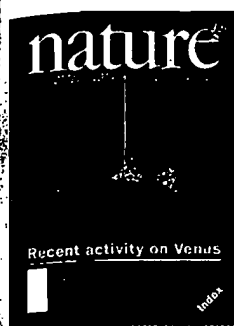
nature

Nature Publishing Company

22/29 December 1994
Vol. 372 Issue no. 6508

◄ A view of Venus based on Magellan radar maps, which covered 98% of the planet's surface; 'gaps' were filled with data from previous US and Soviet missions and the Arecibo radio telescope. Geological features and crater density patterns revealed by such images provide strong evidence for recent volcanic and tectonic activity, suggesting that Venus is probably still active. Pages 756 and 729. (NASA/JPL).

THIS WEEK ... THIS WEEK ... THIS WEEK ...

Author Index
This issue features the complete author and subject index for 1994.

Going down
The relative densities of components of the Earth's crust and mantle are important factors in attempts to predict the dynamics of the Earth's interior. Synthetic MORB (mid-ocean-ridge basalt) glass, subjected to pressures characteristic of the lower mantle, is shown on page 767 to transform to a perovskite-bearing assemblage that should be more dense than the surrounding mantle. This confirms the expectation that if subducted slabs penetrate into the lower mantle, they will be able to sink without hindrance.

Insulin receptor
The crystal structure of the protein tyrosine kinase domain of the human insulin receptor has been determined to 2.1 Å resolution. The overall structure is similar to cAMP-dependent protein kinase, CDK2 and MAP kinase. The structural factors determining tyrosine specificity and its novel autoinhibitory mechanism throw light on the way in which the insulin receptor and other protein tyrosine kinases signal. Pages 746 and 726.

Sun hazard
A study of the induction of skin cancers by sunlight shows that it both initiates and promotes tumours. Initiation occurs when UV light induces mutations in the p53 gene, present in most human skin cancers. But p53 is also responsible for the apoptosis of skin cells with damaged DNA, better known as sunburn, and sunlight can therefore allow the selective expansion of cells with mutated p53. Thus a steep increase in incidence of squamous cell carcinoma of the skin is likely in the next few years in individuals now aged 30–50 who had excessive exposure to sunlight when young. Pages 773 and 730.

Legal matters
The injunction granted in a UK court last month allowing Chiron exclusive rights to market a hepatitis C virus test is good news for pharmaceutical companies. In the third of the series 'Science and the Law', Moss and Cohen argue that the news should be more generally welcomed. Page 814.

Pain centre
The existence of a specific nucleus in the human brain responsible for the perception of pain and temperature sensation was postulated in 1911 on the basis of the characteristics of patients suffering from thalamic pain syndrome. Now Craig *et al.* have used anterograde tracing and immunohistochemical staining to identify this nucleus in the posterior thalamus of man and the macaque monkey, supporting the idea of a central representation of pain. Page 770.

Gadolinium about
A search for the general principles that govern the formation of carbon nanotubes filled with metallic compounds by arc discharge reveals that the most impressive continuous 'nano-wires' are obtained with metals in which an incomplete electronic shell is present in the most stable oxidation state, with chromium and gadolinium

proving the best for the purpose. Above, part of a long (1 μm) nanotube containing crystalline chromium carbide. Distance between parallel fringes of graphitic tubes is 0.34 nm. Pages 761 and 731.

Exhibit 13

Opinion
Genomic analysis of the eukaryotic protein kinase superfamily: a perspective
Steven K Hanks

Address: Department of Cell and Developmental Biology, Vanderbilt University School of Medicine, Nashville, TN 37232, USA.
E-mail: steve.hanks@vanderbilt.edu

Abstract

Protein kinases with a conserved catalytic domain make up one of the largest 'superfamilies' of eukaryotic proteins and play many key roles in biology and disease. Efforts to identify and classify all the members of the eukaryotic protein kinase superfamily have recently culminated in the mining of essentially complete human genome data.

Phosphorylation by protein kinases is recognized as a major mechanism by which virtually every activity of eukaryotic cells is regulated, including proliferation, gene expression, metabolism, motility, membrane transport, and apoptosis. An ultimate goal of research into signal transduction is to reach a full understanding of the protein phosphorylation events that occur within individual cell types and how they eventually impact on cell behavior. A milestone *en route* to this ambitious goal is a determination of the number of protein kinases encoded by eukaryotic genomes and an assessment of their structures, functions, and evolutionary relationships. This article traces the progress made toward achieving these objectives in the pregenomic and genomic eras, which culminated recently with reports on the 'full complement' of human protein kinases.

The pregenomic era
About sixteen years ago, while working at the Salk Institute, my colleagues and I undertook a comparative analysis of all the available sequences of protein kinase catalytic domains [1]. This interest stemmed from my having identified several novel human protein kinases using a homology-based cDNA cloning strategy [2] and wanting to determine their relationships to other known protein kinases. In collaboration with the Salk's resident protein kinase guru Tony Hunter and biocomputing specialist Anne Marie Quinn, we aligned the homologous catalytic-domain amino-acid sequences of 65

distinct protein kinases from diverse eukaryotes (including 45 nonorthologous vertebrate enzymes) and constructed a phylogenetic tree to visualize their overall relationships [1]. The alignment (produced manually at the word-processor) defined the boundaries of the eukaryotic protein kinase (ePK) catalytic domain, revealed conserved subdomains that were never interrupted by amino-acid insertions, and identified highly conserved individual amino acids and motifs (Figure 1).

The phylogenetic tree revealed major clusters including the tyrosine kinases (the TK group), cyclic nucleotide- and calcium-phospholipid-dependent kinases (the AGC group; including the PKA, PKG, and PKC families) and calmodulin-dependent kinases (the CAMK group). These groupings indicated that ePK domain phylogeny reflects substrate specificity and/or mode of regulation and could therefore serve as a useful classification tool. Over the next 7 years I continued to add new sequences to the alignment as they became available and to construct phylogenetic trees as a means of classifying the burgeoning ePK superfamily. By early 1994, the ePK domain alignment had grown to contain 390 sequences including 205 non-orthologous vertebrate ePKs, and a fourth major ePK group (CMGC, comprising the CDK, MAPK, GSK, and CLK families) had been added through phylogenetic analysis [3]. The 390 ePK domain alignment was made publicly available through the Protein Kinase Resource website [4].
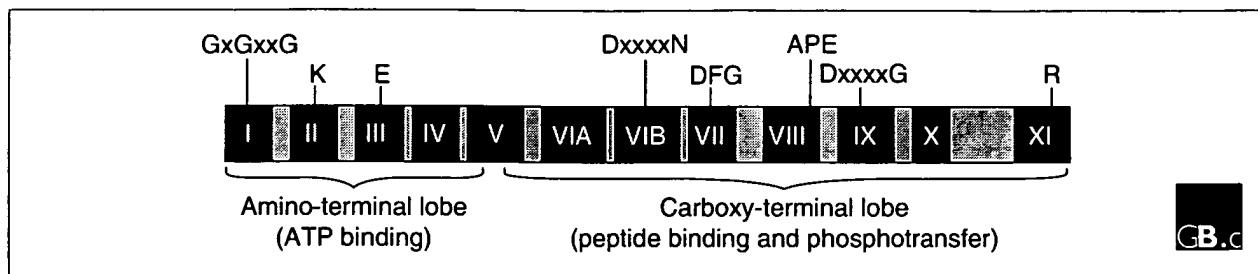
Figure 1
The ePK catalytic domain. The 12 conserved subdomains are indicated by Roman numerals. The positions of amino-acid residues and motifs highly conserved throughout the ePK superfamily are indicated above the subdomains, using the single-letter amino-acid code with x as any amino acid. Crystal structures show that ePK domains adopt a common fold consisting of amino-terminal and carboxy-terminal lobes connected by a hinge region. Binding of Mg-ATP is largely the function of the amino-terminal lobe and hinge region, while peptide-substrate binding is mediated by the carboxy-terminal lobe. Particularly important for catalytic function are the invariant lysine in subdomain II and the invariant aspartate in subdomain VII that function to anchor and orient ATP, and the invariant aspartate in subdomain VIB which is the likely catalytic base in the phosphotransfer reaction. More detailed discussions of ePK subdomains and conserved residues in relation to crystal structures and catalytic function can be found in [3,4,12,13].

The genomic era

By 1995, with the advent of genome-sequencing projects, the task of cataloging and classifying the members of the ePK superfamily had grown to become too distracting from my funded research and I discontinued my efforts in this area. Tony Hunter continued to work with bioinformaticians at SUGEN, Inc. (including Greg Plowman, Gerard Manning, and Sucha Sudarsanam) to characterize the full ePK complements of model eukaryotes from genomic sequence data [5,6]. By the time of a recent report [7], their efforts had resulted in the identification and classification of 115 distinct ePKs from budding yeast (around 2% of all genes), 434 from *Caenorhabditis elegans* (about 2.5% of all genes), and 223 from *Drosophila*. In addition they described the complement of 'atypical protein kinases' (aPKs) from these species: 15 from yeast, 20 from *C. elegans*, and 16 from *Drosophila*. (The aPKs are a variety of protein kinases that lack strong sequence similarity to the classical ePK domain but have been shown experimentally to have protein kinase activity; well-known examples are the 'lipid kinases' of the phosphatidylinositol 3'-kinase (PI3K) family, some of which have been shown experimentally to have protein kinase activity.)

As a result of their comprehensive analyses of 'kinomes', the SUGEN investigators were able to define three new major groups within the broad ePK classification scheme: first, the STE group, which includes ePKs that function in the MAPK kinase cascades that were first described through characterization of yeast *sterile* mutants; second, the CK1 group, including the casein kinase 1 family and related enzymes, which is greatly expanded in the worm; and third, the TKL ('tyrosine-kinase like') group that includes the STKR family of TGFbeta serine/threonine kinase receptors and is phylogenetically close to the tyrosine kinases (TKs). Many distinct kinase families within the AGC, CAMK, CMGC, STE, and CK1 groups have representatives from all three species, supporting the idea of an early evolutionary origin and critical

function in basic cellular processes. Members of the TK and TKL groups are notably absent from yeast, consistent with the known functions of these ePKs in intercellular signaling events associated with metazoan complexity. More discussion of the evolutionary relationships among the ePKs identified through the SUGEN genome-mining efforts has been published elsewhere [7]. The SUGEN kinase.com website [8] includes links to all their published work on protein kinase analysis as well as 'KinBase', a very useful searchable database that holds information on all the protein kinase genes found in the yeast, worm, fly, and human (see below) genomes.

Human protein kinases

The completion of the first draft of the human genome sequence presented an opportunity to determine the full complement of human protein kinases. The first analysis came from a group led by Mitch Kostich at Schering-Plough Research Institute (SPRI) [9]. This group mined public GenBank records (available before December, 2001) for ePK sequences by performing BLAST searches using known ePK domains as queries. The resulting hits were consolidated, and efforts were made to remove non-human sequences, pseudogenes, and poor-quality sequences that could represent duplicate hits. The SPRI investigators chose to err on the side of inclusion rather than exclusion, however, and many cases of 'single hit' sequences were retained. Their effort culminated in a collection of 510 potentially unique human ePKs. A color-coded alignment that accompanied their article [9] nicely illustrates the ePK domain sequence conservation.

The SUGEN group, led by Gerard Manning and Sucha Sudarsanam, carried out a more comprehensive effort to describe and classify all human ePKs [10]. They employed a dataset that included, in addition to the public databases, genomic reads from Celera that are not publicly available, non-public expressed sequence tags (ESTs) from Incyte and SUGEN, and they searched using a hidden Markov model of

the ePK domain that allowed detection of very divergent family members. The sequence data were further searched for members of the various known aPK families. Using stringent criteria to eliminate false positives (including verification of novel sequences by cDNA cloning) they compiled a list of 478 human members of the ePK superfamily and another 40 aPKs, bringing their human kinome total to 518 (approximately 1.7% of all predicted human genes). They also identified 106 ePK or aPK pseudogenes.

A comparison of the SPRI-510 and SUGEN-518 lists reveals 474 protein kinases in common (see the additional file available with this article online). Of the 44 SUGEN-specific kinases, 32 are aPKs; the other 8 aPKs identified by SUGEN, from the ABC1 and RIO families, were included in the SPRI list as a result of their having weak ePK domain similarity. Of the remaining 12 SUGEN-specific ePKs, five (TAK1, MLKL, NEK5, SgK307, and TBCK) were not available in the public data used in the SPRI analysis; another five (SgK196, SgK223, SgK424, SgK493, and Slob) have rather divergent ePK domains that lack many of the highly conserved residues and are unlikely to have catalytic activity, so it is easy to see how these might have been excluded by visual inspection; and the final two are SgK110 and NEK10. SgK110 was actually detected by the SPRI search, but it was erroneously merged with a related sequence AC008735_EPK1 (SgK069) on the same genomic contig; and it is unclear why the SPRI group missed NEK10. Most, if not all, of the 36 SPRI-specific ePKs represent over-inclusion errors (Table 1): 14 correspond to sequences determined to be pseudogenes by the SUGEN group; 19 are based on single sequences that are (or appear to be) either poor-quality duplicates of other ePKs or interspecies contaminants; and the remaining three are duplicates arising by virtue of non-overlapping partial sequences.

Thus the SUGEN compilation of 478 human ePK superfamily genes represents the accurate count based on current sequence data. If one subtracts those that lack key conserved residues, we are left with 428 human ePKs with known or likely kinase function (Table 2), 99% of which were included in the SPRI list; 365 of these fall within the seven major ePK groups: TK, 84 in total; CAMK, 66; AGC, 61; CMGC, 61; STE, 45; TKL, 37; and CK1, 11. The remaining 63 are in the 'Other' category, falling outside the main ePK group branches. Krupa and Srinivasan [11] have also recently searched the public human genome data with a focus on identifying functional protein kinases; their efforts resulted in a list of 448 distinct human ePK sequences, but around 90 of these appear to represent duplicate entries, and no novel protein kinases were identified that were not present in the SUGEN compilation.

Usefulness of the kinome data

Knowing the full complement of ePK family members and functional ePKs encoded by eukaryotic genomes will have great impact upon many areas of scientific investigation. As mentioned above, an obvious benefit relates to understanding

Table 1

Putative ePKs identified by SPRI but not SUGEN

Category	SPRI name	Comment
14 pseudogenes	LOC95530	Corresponds to SUGEN MAP2K2ps
	AC008014_EPK1	Corresponds to SUGEN MARKps22
	AC023095_EPK1	Corresponds to SUGEN PLK1ps1
	AC024933_EPK1	Corresponds to SUGEN MARKps30
	AC091554_EPK1	Corresponds to SUGEN SRPK2ps
	AI621045	Corresponds to SUGEN SgK384ps
	AI991174	Corresponds to SUGEN TSSKps2
	AL138964_EPK1	Corresponds to SUGEN MARKps23
	AL161450_EPK1	Corresponds to SUGEN CK1g2ps
	AL390061_EPK1	Corresponds to SUGEN MNK1ps
	BG742738	Corresponds to SUGEN TLK2ps2
	LOC65348	Corresponds to SUGEN CDK7ps
	LOC65743	Corresponds to SUGEN PAK2ps
	LOC85643	Corresponds to SUGEN CLK2ps
11 duplicates from poor-quality single-EST sequences	AI342911	Corresponds to SUGEN CLK3
	AA565761	Corresponds to SUGEN CLK3
	AI826899	Corresponds to SUGEN CDK4
	AW148845	Corresponds to SUGEN ILK
	AW674733	Corresponds to SUGEN ILK
	AA730234	Corresponds to SUGEN MAP2K7
	AI075923	Corresponds to SUGEN TSSK4
	AI206530	Corresponds to SUGEN TLK1
	AW518044	Corresponds to SUGEN MPSK1
	AL538014	Corresponds to SUGEN BRSK2
	BG986727	Corresponds to SUGEN Obscn
Four non-human sequences	AC025598	*Mus musculus* MAP2K6
	AC026940_EPK1	*Mus musculus* CDK7
	PKCETB	*Rattus norvegicus* PKCeta
	CAMKII	*Sus scrofa* sequence, subsequently recalled
Three duplicates from non-overlapping partial sequences	AC023837_EPK1	Part of EphA6
	AA263006	Part of HIPK1 (along with BE180036)
	BE567816	Part of YANK1 (along with BG036777)
Four others	AW499744	Single EST; does not seem to encode an ePK
	AJ336398_EPK1	Single genomic sequence, possibly prokaryotic
	AL512402_EPK1	Single genomic sequence, possibly nonhuman
	ILK2	Possibly murine ILK; published as human [16]

Table 2

The 428 human ePKs with known or likely kinase catalytic function

Group	Number within group	Family	Number within family	Family members (SUGEN nomenclature)
TK	84	Ack	2	ACK, TNK1
		Abl	2	ABL, ARG
		Csk	2	CSK, CTK
		FAK	2	FAK, PYK2
		Fer	2	FER, FES
		JAK	4	JAK1, JAK2, JAK3, TYK2
		Src	11	BLK, BRK, FGR, FRK, FYN, HCK, LCK, LYN, SRC, SRM, YES
		Syk	2	SYK, ZAP70
		Tec	5	BMX, BTK, ITK, TEC, TXK
		Alk	2	ALK, LTK
		Axl	3	AXL, MER, TYRO3
		DDR	2	DDR1, DDR2
		EGFR	3	EGFR, HER2/ErbB2, HER4/ErbB4
		Eph	12	EphA1, EphA2, EphA3, EphA4, EphA5, EphA6, EphA7, EphA8, EphB1, EphB2, EphB3, EphB4
		FGFR	4	FGFR1, FGFR2, FGFR3, FGFR4
		InsR	3	IGF1R, INSR, IRR
		Lmr	3	LMR1, LMR2, LMR3
		Met	2	MET, RON
		Musk	1	MUSK
		PDGFR/VEGFR	8	FLT3, FMS, KIT, PDGFRa, PDGFRb, FLT1, FLT4, KDR
		Ret	1	RET
		Ror	2	ROR1, ROR2
		Sev	1	ROS
		Tie	2	TIE1, TIE2
		Trk	3	TRKA, TRKB, TRKC
AGC	61	PKA	5	PKACa, PKACb, PKACg, PRKX, PRKY
		PKG	2	PKG1, PKG2
		PKC	9	PKCa, PKCb, PKCd, PKCe, PKCg, PKCh, PKCi, PKCt, PKCz
		AKT	3	AKT1, AKT2, AKT3
		DMPK	7	CRIK, DMPK1, DMPK2, MRCKa, MRCKb, ROCK1, ROCK2
		GRK	7	BARK1, BARK2, GPRK4, GPRK5, GPRK6, GPRK7, RHOK
		MAST	5	MAST1, MAST2, MAST3, MAST4, MASTL
		NDR	4	LATS1, LATS2, NDR1, NDR2
		PKB	1	PDK1
		PKN	3	PKN1, PKN2, PKN3
		RSK	9	MSK1, MSK2, RSK1, RSK2, RSK3, RSK4, SgK494, p70S6K, p70S6Kb
		SGK	3	SGK, SGK2, SGK3
		YANK	3	YANK1, YANK2, YANK3
CAMK	66	CAMK1	5	CaMK1a, CaMK1b, CaMK1d, CaMK1g, CaMK4
		CAMK2	4	CaMK2a, CaMK2b, CaMK2d, CaMK2g
		CAMKL	20	AMPKa1, AMPKa2, BRSK1, BRSK2, CHK1, HUNK, LKB1, MARK1, MARK2, MARK3, MARK4, MEL, NIM1, NuaK1, NuaK2, PASK, QIK, QSK, SIK, SNRK
		DAPK	5	DAPK1, DAPK2, DAPK3, DRAK1, DRAK2
		DCAMKL	3	DCAMKL1, DCAMKL2, DCAMKL3
		MAPKAPK	5	MAPKAPK2, MAPKAPK3, MAPKAPK5, MNK1, MNK2
		MLCK	4	caMLCK, skMLCK, smMLCK, SgK085
		PHK	2	PHKg1, PHKg2
		PIM	3	PIM1, PIM2, PIM3
		PKD	3	PKD1, PKD2, PKD3
		PSK	1	PSKH1
		RAD53	1	CHK2
		Trio	4	Obscn, SPEG, Trad, Trio

Table 2 *(continued)*

Group	Number within group	Family	Number within family	Family members
		TSSK	5	SSTK, TSSK1, TSSK2, TSSK3, TSSK
		CAMK-Unique	1	STK33
CMGC	61	CDK	20	CCRK, CDC2, CDK2, CDK3, CDK4, CDK5, CDK6, CDK7, CDK8, CDK9, CDK10, CDK11, CHED, CRK7, PCTAIRE1, PCTAIRE2, PCTAIRE3, PFTAIRE1, PFTAIRE2, PITSLRE
		MAPK	14	Erk1, Erk2, Erk3, Erk4, Erk5, Erk7, JNK1, JNK2, JNK3, NLK, p38a, p38b, p38d, p38g
		GSK	2	GSK3A, GSK3B
		CLK	4	CLK1, CLK2, CLK3, CLK4
		CDKL	5	CDKL1, CDKL2, CDKL3, CDKL4, CDKL5
		DYRK	10	DYRK1A, DYRK1B, DYRK2, DYRK3, DYRK4, HIPK1, HIPK2, HIPK3, HIPK4, PRP4
		RCK	3	ICK, MAK, MOK
		SRPK	3	MSSK1, SRPK1, SRPK2
STE	45	STE7	7	MAP2K1, MAP2K2, MAP2K3, MAP2K4, MAP2K5, MAP2K6, MAP2K7
		STE20	28	GCK, HPK1, KHS1, KHS2, LOK, MST1, MST2, MST3, MST4, MYO3A, MYO3B, OSR1, PAK1, PAK2, PAK3, PAK4, PAK5, PAK6, SLK, STLK3, TAO1, TAO2, TAO3, YSK1, ZC1/HGK, ZC2/TNIK, ZC3/MINK, ZC4/NRK
		STE11	8	MAP3K1, MAP3K2, MAP3K3, MAP3K4, MAP3K5, MAP3K6, MAP3K7, MAP3K8
		STE-Unique	2	COT, NIK
TKL	37	IRAK	2	IRAK1, IRAK4
		LISK	4	LIMK1, LIMK2, TESK1, TESK2
		LRRK	2	LRRK1, LRRK2
		MLK	9	DLK, HH498, LZK, MLK1, MLK2, MLK3, MLK4, TAK1, ZAK
		RAF	3	ARAF, BRAF, RAF1
		RIPK	5	ANKRD3, RIPK1, RIPK2, RIPK3, SgK288
		STKR	12	ACTR2, ACTR2B, ALK1, ALK2, ALK4, ALK7, BMPR1A, BMPR1B, BMPR2, MISR2, TGFbR1, TGFbR2
CK1	11	CK1	7	CK1a, CK1a2, CK1d, CK1e, CK1g1, CK1g2, CK1g3
		TTBK	2	TTBK1, TTBK2
		VRK	2	VRK1, VRK2
Other	63	Aur	3	AurA, AurB, AurC
		BUB	1	BUB1
		Bud32	1	PRPK
		CAMKK	2	CaMKK1, CaMKK2
		CDC7	1	CDC7
		CK2	2	CK2a1, CK2a2
		IKK	4	IKKa, IKKb, IKKe, TBK1
		IRE	2	IRE1, IRE2
		MOS	1	MOS
		NAK	4	AAK1, BIKE, GAK, MPSK1
		NEK	11	NEK1, NEK2, NEK3, NEK4, NEK5, NEK6, NEK7, NEK8, NEK9, NEK10, NEK11
		NKF1	3	SBK, SgK069, SgK110
		NKF2	1	PINK1
		NKF4	2	CLIK1, CLIK1L
		PEK	4	GCN2, HRI, PEK, PKR
		PLK	4	PLK1, PLK2, PLK3, PLK4
		TLK	2	TLK1, TLK2
		TOPK	1	PBK
		TTK	1	TTK
		ULK	4	Fused, ULK1, ULK2, ULK3
		VPS15	1	PIK3R4
		WEE	3	MYT1, Wee1, Wee1B
		Wnk	4	Wnk1, Wnk2, Wnk3, Wnk4
		Other-Unique	2	KIS, SgK496

Figure 2

	VIB	VIII
TK	H r D l A a ? N / ? t A	1 ? i ? ? m a p E / f v ? t p l
AGC	Y R D l K p e N / H i l d	G T p e y i a P E / d m
CAMK	H r D l k p e N / l i a	g t p x y a a P e / s f v
CMGC	H R D l K p e N / t i s	? T r w Y ? a p e / ? S l y p
STE	H r D i K g x N / v a	G t p y w ? a P E / f y
TKL	h r D l k ? x N	g t x r y m a p E / s a w
CK1	h r D V K p x N / I	G T a r y a ? ? ? / v m

Conserved residues implicated in peptide-substrate recognition. Consensus motifs for the catalytic loop region in subdomain VIB and activation loop region in subdomain VIII were determined for the members of each of the seven major ePK groups with known or likely kinase activity. Invariant residues at a given position are indicated by single upper-case letters. Two upper-case letters at a single position indicate that either of two residues are strictly conserved, the most frequent shown in the top row. Positions in which more than two amino acids are present are indicated with lower-case letters; a single letter indicates that only one residue is highly conserved, two letters indicate that either of two residues are frequently conserved (most frequent on the top row), and 'x' indicating poor positional conservation. Residues highlighted in outline are notably conserved within an ePK group and are thought to function in the recognition of peptide substrates specifically targeted by the members of the group.

of how signal transduction pathways evolved during the course of eukaryotic evolution. Both SUGEN [10] and Krupa and Srinivasan [11] extended their analyses to describe other domains present in the various human ePKs which are likely to function in directing the enzymes to relevant substrates or modulating kinase activities. Further analysis of the ePK domain sequences uniquely conserved within the major groups and families, together with comparisons of ePK domain crystal structures, should ultimately allow a full understanding of how different classes of peptide substrate are recognized. For example, Figure 2 shows consensus sequences for the catalytic loop region in subdomain VIB (which includes the invariant aspartate thought to function as the catalytic base) and the activation loop region in subdomain VIII (which includes the highly conserved glutamine in the 'APE' motif) - two regions that have been recognized as being primarily involved in peptide-substrate recognition [12,13]. A number of group-specific differences are apparent (highlighted in Figure 2) that correlate with unique peptide-recognition tendencies for the ePKs that fall within a given group [14]. Beyond sequence analysis, the kinome data will allow for the development of comprehensive tools (such as full-length cDNAs, microarrays, antibodies, and fusion protein and RNAi constructs) that will greatly aid laboratory investigations aimed at understanding cell signaling through analysis of kinase function. As an example of such proteomic approaches to the study of protein kinases, nearly all yeast protein kinases have been expressed in bacteria and analyzed for their ability to phosphorylate an array of protein or peptide substrates using protein-chip technology [15]. Finally, the human kinome data will have benefits in the understanding and treatment of human diseases. The ePK genes that map within disease loci are attractive etiological candidates, and knowledge of the full repertoire of human protein kinases will greatly aid in the development of drugs that target specific protein kinases or protein kinase families whose function contributes to disease-associated cellular defects.

Acknowledgements

Additional data file

An additional data file with the 474 protein kinases in common between the SPRI-510 and the SUGEN-518 lists is available with the online version of this article.

References

1.   Hanks SK, Quinn AM, Hunter T: **The protein kinase family: conserved features and deduced phylogeny of the catalytic domains.** *Science* 1988, 241:42-52.
2.   Hanks SK: **Homology probing: identification of cDNA clones encoding members of the protein-serine kinase family.** *Proc Natl Acad Sci USA* 1987, 84:388-392.
3.   Hanks SK, Hunter T: **The eukaryotic protein kinase superfamily: kinase (catalytic) domain structure and classification.** *FASEB J* 1995, 9:576-596.
4.   **Protein Kinase Resource** [http://kinases.sdsc.edu/html/index.shtml]
5.   Hunter T, Plowman GD: **The protein kinases of budding yeast: six score and more.** *Trends Biochem Sci* 1997, 22:18-22.
6.   Plowman GD, Sudarsanam S, Bingham J, Whyte D, Hunter T: **The protein kinases of *Caenorhabditis elegans*: a model for signal transduction in multicellular organisms.** *Proc Natl Acad Sci USA* 1999, 96:13603-13610.
7.   Manning G, Plowman GD, Hunter T, Sudarsanam S: **Evolution of protein kinase signaling from yeast to man.** *Trends Biochem Sci* 2002, 27:514-520.
8.   **Kinase.com** [http://kinase.com/]
9.   Kostich M, English J, Madison V, Gheyas F, Wang L, Qiu P, Greene J, Laz TM: **Human members of the eukaryotic protein kinase family.** *Genome Biol* 2002, 3:research0043.1-12.

Exhibit 14

REVIEWS

The protein kinases of budding yeast: six score and more

Tony Hunter and Gregory D. Plowman

The completion of the budding yeast genome sequencing project has made it possible to determine not only the total number of genes, but also the exact number of genes of a particular type[1-3]. As a consequence, we now know exactly how many protein kinases are encoded by the yeast genome, a number of considerable interest because of the importance of protein phosphorylation in the control of so many cellular processes.

BUDDING YEAST has 113 conventional protein kinase genes, corresponding to ~2% of the total genes (see Table I in centrefold). More than 60% of these protein kinases have either known or suspected functions; the remainder are novel, and functional analysis awaits. In terms of defined functions encoded by the yeast genome, protein kinases come a close second behind transcription factors[3].

What can be learnt from knowing all the protein kinases encoded by a single eukaryotic genome? One obvious outcome is that it is possible to say whether a protein kinase identified and characterized in another organism has a homologue in budding yeast. Such a homologue might either have a known function in yeast, or its function can be tested by genetic studies. Equally important is the recognition of protein kinase subfamilies present in higher eukaryotes that are absent from yeast. Whereas all eukaryotes have similar requirements for DNA replication, transcription, translation and energy metabolism, it is reasonable to expect that there might be many protein kinases unique to multicellular organisms that function in cellular communication, both between cells, tissues and the environment, such as protein-tyrosine kinases. Conversely, there might be protein kinases unique to budding yeast.

T. Hunter is at The Salk Institute, 10010 North Torrey Pines Road, La Jolla, CA 92037, USA. Email: hunter@salk.edu G. D. Plowman is at SUGEN, Inc., 515 Galveston Drive, Redwood City, CA 94063, USA. Email: plowman@sugen.sf.ca.us

Multiple alignment and parsimony analysis of catalytic domain sequences (Fig. 1) categorizes the yeast protein kinases into subfamilies based on structural relatedness. From such a classification, one can infer functional similarities, including regulation of catalytic activity, substrate specificity and cellular localization. This information is of particular value for understanding the function of the numerous uncharacterized yeast open reading frames that exhibit protein kinase motifs. One can also determine which protein kinases subfamilies are conserved or expanded in other organisms and which are unique to yeast.

MAP kinase pathways

Pseudohyphal development. One of the virtues of knowing all yeast protein kinases is that it delimits the number of protein kinases in a particular subfamily. For example, five members of the mitogen-activated protein (MAP) kinase family, Fus3, Kss1, Hog1, Mpk1/Slt2 and Smk1, had been identified and functionally characterized before the completion of the genome project[4]. However, both haploid invasive growth and diploid pseudohyphal development of budding yeast are known to require Ste11 and Ste7, which are a MAP kinase kinase kinase (MAPKKK) and a MAP kinase kinase (MAPKK), respectively. These kinases normally function in the mating pheromone response pathway to activate the Fus3 (or Kss1) MAP kinase, and also Ste12, a transcription factor that is phosphorylated and activated by Fus3 (Ref. 5). This suggested that a MAP kinase would be required for pseudohyphal development, but none of the five characterized MAP kinases proved to be essential for haploid invasive

growth, either singly or in combination (H. Madhani and G. Fink, pers. commun.).

The putative protein kinase YKL161C clusters with the MAP kinases (Fig. 1) and might be another MAP kinase (although it has a KxY rather than TxY motif in the activation loop), but similarly, this gene is not essential for pseudohyphal development. As Ste7 normally phosphorylates the Thr and Tyr in the TxY motif in the activation loop of Fus3 and Kss1, one possibility is that Ste7 might phosphorylate and activate another protein kinase with a related activation-loop motif. Three other protein kinases, Kin3/Npk1, Ssn3/Srb10 and Ime2, have a TxY activation-loop sequence and might, in principle, be regulated by Ste7 phosphorylation. Ssn3/Srb10 is a cyclin-dependent kinase (CDK) and is unlikely to be a Ste7 target. Ime2, however, has motifs characteristic of proline-directed protein kinases, like the MAP kinases, and is on the same major branch as the MAP kinases (Fig. 1), making it a potential candidate, although Ime2 is normally only produced during meiosis[6].

With regard to the protein kinases lying upstream of the MAP kinases, there are four MAPKKs (Ste7, Mkk1, Mkk2 and Pbs2: STE7/MEK family), and four MAPKKKs (Ste11, Bck1, Ssk2 and Ssk22: STE11/MEKK family). These have been assigned to the Fus3+Kss1, Mpk1 or Hog1 MAP kinase pathways by genetical and biochemical analyses[4]. This leaves both the Smk1 MAP kinase, which is required for spore wall assembly[7], and YKL161C without their own specific MAPKK and MAPKKK. Possibly, a MAPKKK/MAPKK combination used for one of the other MAP kinase pathways is also used for these predicted MAP kinases. Interestingly, however, the Smk1 pathway might have a specific MAPKKK kinase, as the Sps1 protein kinase (NRK/MESS family), which lies upstream of Smk1 (Ref. 8), is related to Ste20, a MAPKKK kinase in the Fus3 MAP kinase pathway.

The Ste20 family. Ste20 is a Cdc42-activated protein kinase[9] that is required in the pheromone response MAP kinase pathway upstream of Ste11 (Refs 10, 11). A *ste20* mutant is viable and only shows a defect in response to mating phero-mone[11]. Cla4 is a second member of the Ste20 family, which also interacts with Cdc42, and has been implicated in polarized cell growth, budding and cytokinesis, but a *cla4* mutant is viable despite a cytokinesis defect[12].

A double *ste20/cla4* mutant, however, cannot undergo cytokinesis, implying

Table I. Classification of *Saccharomyces cerevisiae* protein kinases (contd)

NIMA/NEK family (Similar to NIMA_en, NIMA1_h, NEK1_h)
  KIN3/NPK1/FUN52/YAR018C Ser/Thr protein kinase; null mutation has no phenotype. TTY in kinase subdomain VIII (activation loop)

NEK-like family (Similar to F35G12.3_ce, weakly to NEK1_h)
  YNL020C/N2823 Ser/Thr protein kinase of unknown function
  YIL095W/(PAK1) Ser/Thr protein kinase of unknown function
  YBR059C/YBR0419 Ser/Thr protein kinase of unknown function

Other group (24 members)

Casein kinase I family (Similar to CKI_h)
  YCK1/CKI2/YHR135C Casein kinase I (CKI) isoform
  YCK2/(CKI1)/(CKI)/N1755/YNL154C CKI isoform
  YCK3/CKI3/YER123W CKI isoform
  HRR25/P1850/YPL204W CKI, Ser/Thr/Tyr protein kinase; associated with DNA repair and meiosis

Casein kinase II family (Similar to CKA2_h)
  CKA1/YIL035C CKII, catalytic (α) subunit
  CKA2/02810/YOR061W CKII, catalytic (α) subunit
  *CDC7/SAS1/0AF2/D2855/YDL017W Protein kinase required for initiation of DNA synthesis, for commitment to sporulation, for DNA repair and for meiotic recombination. (Similar to HSK1_sp)

NPR/HAL5 family (Unique to *S. cerevisiae*)
  HAL5/J0531/YJL165C Ser/Thr protein kinase involved in salt and pH tolerance
  YKL168C/YKL632 Ser/Thr protein kinase of unknown function
  SAT4/YCR101/YCR046/YCR008W Protein with similarity to Npr1p protein kinase
  YJR059W/J1725 Putative Ser/Thr protein kinase of unknown function
  PTK1/YKL198C Ser/Thr protein kinase that enhances spermine uptake. (Frame shift corrected)
  NPR1/N1831/YNL183C Ser/Thr protein kinase involved in regulating transport systems for nitrogen nutrients under conditions of nitrogen catabolite derepression
  YDL214C/D1014 Ser/Thr protein kinase with similarity to Npr1
  YDL025C/D2810 Protein with similarity to protein kinase Npr1p
  YOR267C/03420 Ser/Thr kinase protein kinase with similarity to Npr1p

ELM family (Unique to yeast, similar to D45882_sp)
  PAK1/SYGP-ORF45/YER129W Protein kinase capable of suppressing DNA polymerase α mutations
  YGL179C/BIE860/G1618 Ser/Thr protein kinase with similarity to Elm1p and Kin82p
  ELM1/YKL261/YKL048C Ser/Thr protein kinase regulating pseudohyphal development

RAN family (Similar to RAN_sp, p78_h)
  SHA3/SKS1/LPB5/YPL026C Ser/Thr protein kinase; suppressor of hta1 mutations that cause aberrant transcription
  YDR247W/YD8419.14 Ser/Thr protein kinase with similarity to *S. pombe* RAN. negative regulator of sexual conjugation and meiosis
  *KSP1/YHR082C Ser/Thr kinase with similarity to CKII

PIM-like family (Similar to PIM2_m, KIAA0135_h)
  *YAL017W/YAL002/FUN31 Ser/Thr protein kinase of unknown function
  *YOL045W/02034/YOL044W Ser/Thr protein kinase of unknown function

Unique kinases (17 members) (No similar *S. cerevisiae* kinases)

Kinases with possible homologues in other species
  *CDC5/PKX2/M8D2/YM8270.03/YMR001C Ser/Thr protein kinase required for exit from mitosis; ts mutants block after nuclear division. (Similar to PLK-1_h, POLO_dm)
  *IPL1/P1820/YPL209C Ser/Thr protein kinase involved in chromosome segregation. (Similar to AUR_dm)
  *IRE1/ERN1/YHR079C Protein kinase and type I membrane protein involved in signal transduction from ER lumen to nucleus; part of the unfolded protein response. (Similar to C41C4.4_ce)
  *VPS15/VPT15/(VPL19)/YBR0825/YBR097W Ser/Thr protein kinase involved in vacuolar protein sorting. (Similar to ZK930.1_ce)
  *YPL236C/P1057 Protein of unknown function. No GxG (Similar to C3H1.13_sp) (Not in YPD)
  *YGL180W/G1815 Ser/Thr protein kinase of unknown function. (Similar to UNC-51_ce, PLO1_sp)
  *SWE1/J040C/YJL187C Ser/Tyr dual-specificity protein kinase able to phosphorylate Cdc28p on tyrosine and inhibit its activity. (Similar to Wee1_sp and MLK_h)
  *SPK1/RAD53/MEC2/SAD1/P2588/YPL153C Ser/Thr/Tyr protein kinase with a checkpoint function in S and G2. Contains FHA domain. (Similar to CDS1_sp)
  *MPS1/RPK1/D2785/YDL028C Ser/Thr/Tyr protein kinase involved in spindle pole body duplication (Similar to ESK_m, TTK_h)
  *YKL116C/YKL518 Ser/Thr protein kinase with similarity to *S. pombe* NIM1 protein kinase. (33% identity to p78_h)

Table I. Classification of *Saccharomyces cerevisiae* protein kinases (contd)

Kinases with possible homologues in other species contd

*GCN2/AAS1/D9954.16/YDR283C Ser/Thr protein kinase that regulates initiation of translation by phosphorylation of eIF2α (Sui2p) (Similar to EIF2aK_r, HRI_r)

*YBR274W/YBR1742 Protein kinase with similarity to members of the growth factor and cytokine receptor family. (Similar to CHK1-sp, CHK1_ce, SNF1_r)

*YGR262C/G9334 Protein with similarity to apple tree CaM-binding protein kinase PIR:JQ2251. Lacks GxG – not in alignment. (Similar to O-sialoglycoprotein endopeptidase from *Methanococcus jannaschii*)

Kinases without known homologue

*BUB1/G7542/YGR188C Ser/Thr protein kinase and checkpoint protein required for cell-cycle arrest in response to loss of microtubule function. (Amino terminus similar to MAD3_sc)

*YKL171W/YKL635 Ser/Thr protein kinase of unknown function

*YGR052W/G4329 Protein of unknown function

*YPR106W/P8283.9 Protein with similarity to protein kinases Gcn2p; galactosyltransferase-associated protein kinase P58/Gtap, and the Raf proto-oncoprotein

---

Atypical protein kinases (1 member)

*YGR080W/G4583 Protein with similarity to human tyrosine kinase A6 PIR:A55922.

Miscellaneous kinases

Phosphatidylinositol-kinases

TOR1/DRR1/J1803/YJR066W Phosphatidylinositol kinase (PI kinase) homologue involved in cell growth and sensitivity to the immunosuppressant rapamycin

TOR2/DRR2/YKL203C PI kinase homologue involved in cell growth and sensitivity to the immunosuppressant rapam_cin, similar to Tor1p

VPS34/VPT29/(VPL7)/END12/L9672.10/YKR240W PI 3-kinase required for vacuolar protein sorting; activated by protein kinase Vps15p

PIK1/PIK41/PIK120/N0795/YNL267W PI 4-kinase; generates PtdIns(4)P

STT4/L2142.4/YLR305C PI 4-kinase; mutants are staurosporine-sensitive and suppressible by overproduction of Pkc1p

MEC1/ESR1/SAD3/YBR1012/YBR136W Checkpoint protein required for mitotic growth; DNA repair and mitotic recombination (PI kinase homologue)

TEL1/YBL0706/YBL088C Protein involved in controlling telomere length; might have PI 3-kinase or protein kinase activity

YHR099W Protein with weak similarity to Tor1p and Tor2p; possible PI kinase homologue

FAB1/YFR019W Probable PtdIns(4)P 5-kinase involved in orientation or separation of mitotic chromosomes

MSS4/YD8142A.05/YD8142.05 Potential PtdIns(4)P 5-kinase; multicopy suppressor of *stt4* mutation

Guanylate kinases

GUK1/D9461.39/YDR454C Guanylate kinase

Histidine protein kinases

SLN1/YPD2/YIL147C Two-component signal transducer with both a His kinase domain and a receiver domain that functions in the high osmolarity signal transduction pathway

YIL042C Related to the mitochondrial branched-chain α-ketoacid (BKCD) and pyruvate dehydrogenase (PDH) kinases, which are protein-serine kinases

Misclassified kinases

MSS2/D2340/YDL107W/ORF2340 Involved in the expression of mitochondrial cytochrome C oxidase subunit 2 (COX2). (Ser/Thr protein kinase that suppresses the growth defect of *snf3* mutants on low glucose.)

YDR109C/YD9727.05 Protein with similarity to FGGY protein kinase family

YJR061W Putative Ser/Thr protein kinase of unknown function. (Similar to YKL200c_sc, YKL201c_sc)

YLR063W/L2174 Ser/Thr protein kinase of unknown function (Unique)

YML059C/YM9958.03 Protein kinase of unknown function (Similar to ZK370.4_ce, M110.7_ce)

YMR192W/YM9848.04 Ser/Thr protein kinase of unknown function. (Leucine repeat and possible coiled coil. Similar to YPL249C_sc, U49940_ce)

YGL059W/G3441 Protein with similarity to rat branched-chain α-ketoacid dehydrogenase kinase PIR:U27456

YGL227W/G0958 Protein with similarity to *Dictyostelium discoideum* non-receptor tyrosine kinase U32174. 37% identity over 64 residues in amino-terminal non-catalytic domain

YOR287C/O5492 Protein with weak similarity to PITSLRE protein kinase isoforms

---

ᵃThe budding yeast protein kinases are subdivided into distinct families based on structural similarity in their catalytic domains. The overall classification is based on that devised by Hanks and Hunter[38]. Individual kinases are listed by their preferred gene name as established by the Saccharomyces Genome Database at Stanford, followed by their synonyms and a brief description as maintained in the Yeast Protein Database (YPD) available on the Internet at http://www.proteome.com. Additional notes, sequence corrections or close homologues are in parentheses following each entry. Protein kinases preceded by an asterisk share only weak similarity to other members. Entries listed as misclassified kinases are flagged as protein kinases in YPD, but have no structural similarity to the protein kinase family and were excluded from this analysis (some of these represent non-catalytic regulatory domains). Three additional open reading frames were identified that encode protein kinases that were not present in the YPD listing: YPL150W (Group IIB), YMR216C (Group IIID) and YPL236C (Group VII). These new protein kinases were recognized following a comprehensive analysis of the complete yeast DNA sequence using the MPSRCH software (Oxford Molecular) implementation of the Smith Waterman algorithm on a Maspar parallel computer.

Table I. Classification of *Saccharomyces cerevisiae* protein kinases[a]

AGC group (17 members)

PKA family: cAMP-dependent (Similar to PKA_h)
TPK1/PKA1/SRA3/PK25/J0541/YJL164C PKA 1, catalytic subunit
TPK2/PKA2/YKR1/P1855/YPL203W PKA 2, catalytic subunit
TPK3/PKA3/YKL630/YKL166C PKA 3, catalytic subunit

PKC family: DAB-activated, PL-dependent (Similar to PKC_h)
PKC1/STT1/HPO2/CLY15/YBL0807/YBL105C PKC; regulates MAP kinase cascade involved in regulating cell wall metabolism

AGC family (Similar to SCK1_sp, RACa_h, AKT_h – no diacylglycerol-binding domain)
SCH9/KOM1/YHR205W Ser/Thr protein kinase activated by cAMP; overproduction can suppress *cdc25* mutant
YPK1/YKL126W Ser/Thr protein kinase with similarity to PKC
YPK2/YKR2/YM9718.03/YMR104C Ser/Thr protein kinase with similarity to Ypk1p

S6K (70 kDa) family (Similar to KAD5_sp, SGK_r)
KIN82/YCR1153/YCR091W Ser/Thr protein kinase of unknown function
YNR047W/N3449 Ser/Thr protein kinase of unknown function

DBF2 family (Similar to KAIB_sp, NDR_h)
DBF2/G4643/YGR092W Ser/Thr protein kinase similar to Dbf20p; required for anaphase/telophase
DBF20/P6263.6/YPR111W Cell-cycle protein kinase similar to Dbf2p; involved in M-phase termination

PKA-related family (Unique to *S. cerevisiae*)
YOL100W/HRC1081/O0784 Ser/Thr protein kinase of unknown function
YDR490C/D6035.33 Ser/Thr protein kinase of unknown function
YDR466W/D6035.10 Ser/Thr protein kinase of unknown function

Other AGC family (Similar to CEK1_sp, COT1_nc, NDR_h, LATS_dm, MAST205_m)
YFL033C Ser/Thr protein kinase with similarity to *Schizosaccharomyces pombe* CEK1 protein kinase
YNL161W/N1727 Putative Ser/Thr protein kinase of unknown function
*YBR028C/YBR0312 Ser/Thr protein kinase with similarity to Ypk2p/Ykr2p and Ypk1p

---

CaMK group (18 members)

CaMK family: Ca^{2+}-calmodulin regulated (Similar to CaMKI_h)
CMK1/YFR014C Ca^{2+}-calmodulin-dependent Ser/Thr protein kinase (CaM kinase), type I
CMK2/O2325/YOL016C CaM kinase type II
RCK1/G1854/YGL158W Ser/Thr protein kinase with similarity to Cmk1p, Cmk2p and Cmk3p. (Sequence updated)
RCK2/CMK3/CLK1/L9672.6/YLR248W CaM kinase

SNF1/AMPK family (Similar to AMPK_h, NPK5_nt, PAR1_ce)
SNF1/CAT1/CCR1/PAS14/HAF3/D8035.20/YDR477W Ser/Thr protein kinase essential for derepression of glucose-repressed genes; acts with Snf4p
KIN1/YDR737.17/YDR122W Ser/Thr protein kinase; similar to Kin2p and *S. pombe* KIN1
KIN2/L9004.3/L2346/YLR096W Ser/Thr protein kinase; similar to Kin1p and *S. pombe* KIN1
KIN4/KIN31/(KIN3)/O5220/YOR233W Ser/Thr protein kinase; similar to Kin1p and Kin2p; catalytic domain is most similar to Snf1p
YPL141C/LP15 Ser/Thr protein kinase with similarity to Kin4p
YPL150W/P2697 Ser/Thr protein kinase of unknown function. (Not in YPD listing)

GIN4 family (Similar to SNF1_sc)
GIN4/D9719.13/YDR507C Ser/Thr protein kinase with similarity to Yc1024p; growth inhibitory protein
YCL024W Protein with similarity to Snf1p
HSL1/YKL453/YKL101W Ser/Thr protein kinase that interacts genetically with histone mutations

Other CaMK family (Similar to Z71478_sp, MLCK_dd, CaMKI_h)
MEK1/MRE4/O6357/YOR351C Ser/Thr protein kinase required for meiotic recombination; contains forkhead-associated (FHA) nuclear signalling domain
DUN1/ORF2370/YDL101C Protein kinase necessary for induction of Rnr3p and DNA repair genes after DNA damage; contains FHA domain
*YMR291W/ORF830348 Ser/Thr protein kinase of unknown function

---

CMGC group (21 members)

CDK family (Similar to CDK2_h)
CDC28/CDK1/SRM5/HSL5/YBR1211/YBR160W Cyclin-dependent protein kinase (CDK) essential for completion of START and for mitosis; associates with Cks1p and cyclins. PSTAIRE in kinase subdomain III
PHO85/P7102_18A/YPL031C CDK that interacts with cyclin Pho80p to regulate phosphate pathway. PSTAIRE in kinase subdomain III

Table I. Classification of *Saccharomyces cerevisiae* protein kinases (contd)

CDK family contd

CAK1/CIV1/YFL029C CDK-activating kinase (Ser/Thr protein kinase) responsible for *in vivo* activation of Cdc28p. PHNAKFE in kinase subdomain III

SSN3/UME5/SRB10/(ARE1)/P7102_08/YPL042C Ser/Thr CDK of the RNA polymerase II holoenzyme complex and mediator (SRB) subcomplex. SQSACRE in kinase subdomain III. TLY in kinase subomain VIII

KIN28/ORF2330/YDL108W Ser/Thr CDK component of transcription initiation factor TFIIH; phosphorylates carboxy-terminal domain (CTD) of RNA polymerase large subunit. DMSAIRE in kinase subdomain III

MAPK family (Similar to SPK1_sp, ERK_h)

KSS1/G4149/YGR040W Ser/Thr protein kinase; redundant with Fus3p for induction of mating-specific genes by mating pheromone. TEY in kinase subdomain VIII (activation loop)

FUS3/DAC2/YBL0303/YBL03.21/YBL016W Ser/Thr protein kinase required for cell-cycle arrest and for cell fusion during mating. TEY in activation loop

HOG1/SSK3/L9354.2/L2931/YLR113W Ser/Thr protein kinase; involved in high-osmolarity signal transduction pathway. TGY in activation loop

SLT2/MPK1/SLK2/BYC2/YHR030C Ser/Thr protein kinase involved in the cell wall integrity pathway. TEY in activation loop

YKL161C/YKL615 Ser/Thr protein kinase of unknown function. KGY in activation loop

SMK1/YP9499.10/YPR049W Sporulation-specific MAP kinase required for completion of sporulation. TNY in activation loop

GSK3 family (Similar to SHAGGY_dm, GSK3_h)

MCK1/(YPK1)/N0392/YNL307C Ser/Thr/Tyr protein kinase (meiosis and centromere regulatory kinase); positive regulator of meiosis and spore formation

YOL128C/O0530/ORF1209713 Ser/Thr protein kinase of unknown function

MDS1/RIM11/GSK3/YM9375.08/YMR139W Ser/Thr protein kinase; homologue of mammalian GSK3

MRK1/D2459/D2461/YDL079C Ser/Thr protein kinase with similarity to Mds1p

CLK family

KNS1/L1224/YLL019C Ser/Thr protein kinase of unknown function. (Similar to CLK_h)

YAK1/YJL141C Ser/Thr protein kinase that suppresses loss of Tpk1p + Tpk2p + Tpk3p. (Similar to KA23_sp, MNB_dm, MNB_h)

YMR216C/YM8261.10 Putative Ser/Thr protein kinase; has similarity to Cdc31p. (Similar to DSK1_sp, SRPK1_ce, U52111_h). (Not in YPD listing)

IME2/SME1/J0817/YJL106W Ser/Thr protein kinase and positive regulator of sporulation genes essential for initiation of meiosis. TAY in activation loop. (Similar to MAK_r, p34_h, CDK2_h)

Other CMGC family (Similar to PITSLRE_h, CHED_h)

SGV1/BUR1/P9584.8/YPR161C Ser/Thr protein kinase involved in pheromone adaptation pathway and in cell cycle. PITAQRE in kinase subdomain III

CTK1/YKL139W CTD kinase a subunit; CDK that phosphorylates CTD of RNA polymerase II large subunit. PITSIRE in kinase subdomain III

STE11/STE20 group (10 members)

STE11/MEKK family (Similar to BYR2_sp, NPK1_nt, MEKK_h)

STE11/L8039.10/YLR362W Ser/Thr protein kinase; component of the pheromone pathway and a pathway regulating pseudohyphal development

BCK1/(SLK1)/SSP31/LAS3/SAP3/J0906/YJL095W Ser/Thr protein kinase; involved in the cell wall integrity pathway

SSK2/N3276/YNR031C MAP kinase kinase kinase (MEKK) of the high osmolarity signal transduction pathway

SSK22/YCR073C MEKK with strong similarity to Ssk2p; participates in the high osmolarity signal transduction pathway

STE20/PAK family (Similar to PAK_dm, PAK1_h, PAK65_h, RAC_h)

STE20/YHL007C Ser/Thr protein kinase in the pheromone pathway; also participates in pathway regulating pseudohyphal development

CLA4/ERC10/N0450/YNL0450/YNL298W Ser/Thr protein kinase required for cytokinesis; has similarity to Ste20p

YOL113W/HRA655/O0722 Ser/Thr protein kinase with similarity to Ste20p

NRK/MESS family (Similar to MESS1_m, ZC504.4_ce)

NRK1/KIC1/H8263.14/YHR102W Ser/Thr protein kinase that interacts with Cdc31p

SPS1/D9719.27/YDR523C Ser/Thr protein kinase involved in middle/late stage of meiosis

*CDC15/YAR019C Protein kinase of the MAP kinase kinase kinase family essential for late nuclear division. (Similar to MESS1_m, CDC7_sp, MST1_h)

STE7/MEK group (8 members)

STE7 family (Similar to HST7_ce, MEK1_h)

STE7/D1525/YDL159W Ser/Thr/Tyr protein kinase of MAP kinase kinase (MEK) family; component of the pheromone pathway and a pathway regulating pseudohyphal development

PBS2/H0O4/SFS4/SSK4/OSR1/J0699/YJL128C Ser/Thr/Tyr protein kinase of the MEK kinase family; essential component of the high-osmolarity signal transduction pathway

MKK1/SSP32/O5095/YOR231W Ser/Thr/Tyr protein kinase of the MEK family involved in cell wall integrity pathway

MKK2/SSP33/LPI6/YPL140C Ser/Thr/Tyr protein kinase of the MEK family involved in cell wall integrity pathway. (Sequence updated)
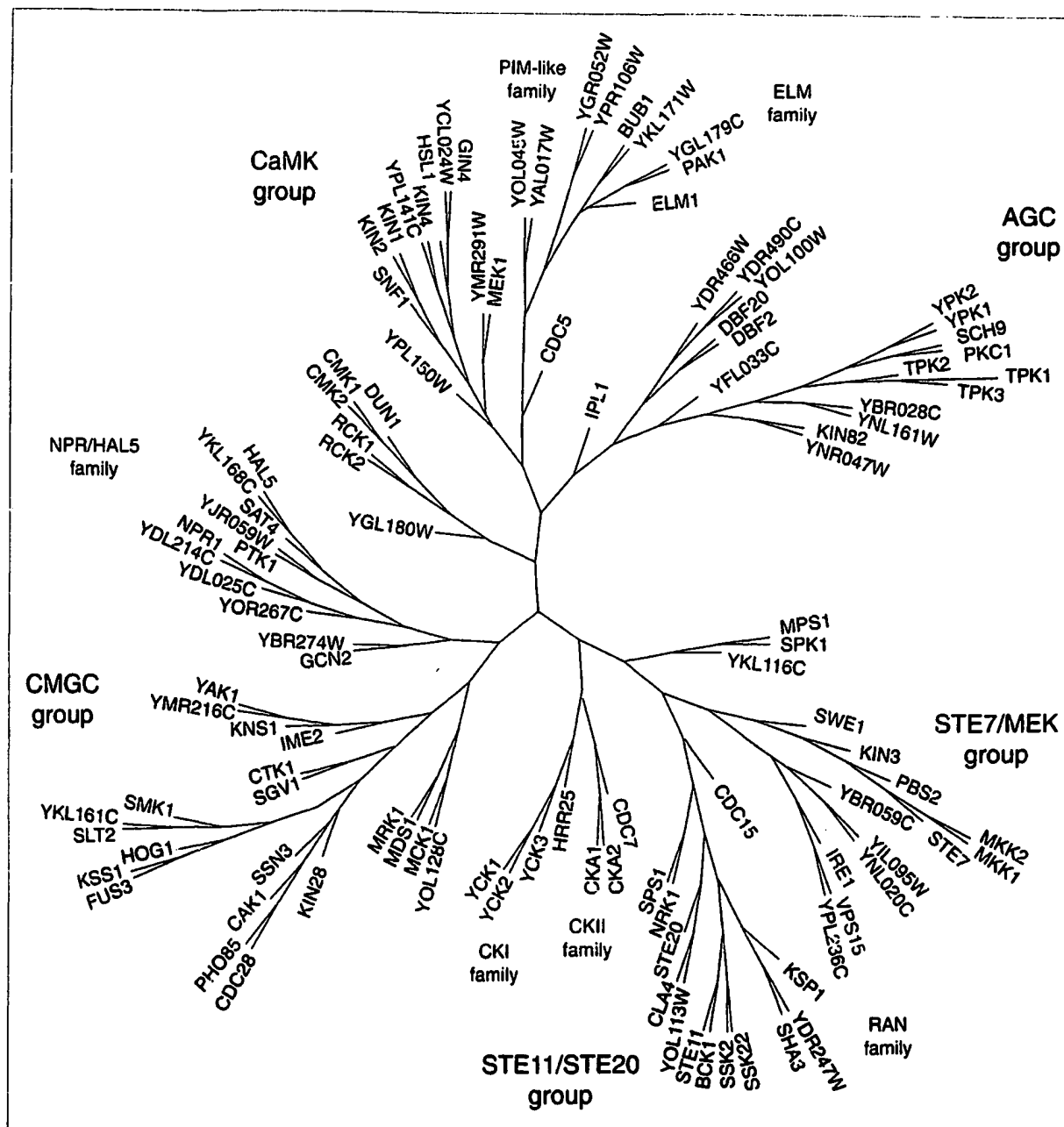
Figure 1

Dendrogram of the budding yeast protein kinase superfamily. The catalytic domains of 113 yeast protein kinases were aligned using SAM[39], a multiple sequence alignment program that applies a linear hidden Markov model to facilitate recognition of conserved subdomains within a protein family (http://www.cse.ucsc.edu/research/compbio/sam.html). The SAM alignment was run on a MasPar parallel computer and the results were inputted into PROTPARS, a protein sequence parsimony method, to build an unrooted phylogeny. PROTPARS is part of the PHYLIP package written by J. Felsenstein of the University of Washington (http://evolution.genetics.washington.edu/phylip.html).

that Ste20 and Cla4 share a function in this process. Nevertheless, the double *ste20/cla4* mutant can still assemble actin, a process regulated by Cdc42 and other Rho family members. This could be accounted for by YOL113W, a third member of the Ste20 family whose existence has been revealed by the genome sequencing project; it will obviously be important to analyse the function of YOL113W and determine to what extent its function is redundant with those of Ste20 and Cla4. Like the other two family members, YOL113W has both a phospholipid-binding pleckstrin homology (PH) domain and a Rac/Cdc42-binding motif (SxPx$_{4-6}$HxxH) upstream of its catalytic domain.

The function of Ste20 subfamily protein kinases is of significant interest, because it expanded greatly during evolution and there are a large number of Ste20-related protein kinases in vertebrates, such as the Paks, which are regulated by Cdc42 and other members of the Cdc42 GTPase family.

Cell-cycle control

The regulation of the cell cycle involves many types of protein kinase. In

budding yeast, Cdc28 is the only known CDK with an essential role in cell-cycle regulation, although two other protein kinases, Pho85 and Kin28, bind and are activated by cyclins and might have roles in cell-cycle progression[13]. Pho85 can play a direct role in G1 regulation, whereas Kin28 has an indirect role as a component of the basal transcription factor TFIIH. All three of these yeast protein kinases contain a canonical AIRE sequence in catalytic subdomain III, which is part of the α1-helix that interacts through the conserved Ile with the cyclin subunit.

An additional member of the CDK cluster, Ssn3/Srb10, has no apparent role in cell-cycle regulation, but forms a complex with cyclin Srb11 as part of the larger RNA polymerase II holoenzyme[13]. Ctk1 is a divergent CDK identified as an RNA polymerase II carboxy-terminal domain kinase, which binds Ctk2, a cyclin-related protein[13], and contains SIRE instead of AIRE. The genome sequence revealed a number of new cyclins, but only one new CDK-related gene, CAK1/CIV1/YFL029C. This gene is unlikely to encode a true CDK, because it has an AKFE sequence instead of the AIRE motif; indeed CAK1/CIV1/YFL029C has recently been shown to be active as a monomer and to be a CDK-activating kinase (CAK) that phosphorylates Thr in the activation loop of Cdc28 (Refs 14–16), a function that in mammalian cells is currently thought to be carried out by cyclin H/Cdk7. The CDK subfamily evidently underwent significant expansion during the evolution of the multicellular eukaroytes, as in vertebrates at least four CDKs are directly involved in cell-cycle regulation and eight CDKs are known altogether.

Another cell-cycle regulatory protein kinase, NIMA, which is required for the G2–M transition in *Aspergillus*, has so far been only identified in filamentous fungi[17]. Given the highly conserved nature of cell-cycle regulation, one might anticipate that other eukaryotes would have NIMA homologues, and some evidence for a NIMA response pathway in vertebrates has been obtained[18,19]. However, no true homologues have been identified in other eukaryotes, although there are several NIMA-related protein kinases in vertebrates. The yeast Kin3/Npk1 protein kinase is quite closely related to NIMA in its catalytic domain[20,21], but lacks the carboxy-terminal regulatory domain, which is critical for NIMA cell-cycle function. Thus, it appears that there might be no budding yeast homologue of NIMA.

There are several solitary yeast cell-cycle regulatory protein kinases, including Cdc5, Cdc7 and Cdc15. Cdc5 is required for exit from mitosis[22], and its three counterparts in higher vertebrates, Plk, Fnk/Prk and Sak, form a subfamily. Plk, Polo (a *Drosophila* relative) and Plo1 (a fission yeast relative), like Cdc5, have been implicated in progression through mitosis. However, Cdc7, a protein kinase in the casein kinase II family, which is required for initiation of DNA synthesis during S phase[23] and is related to Hsk1 from fission yeast, has no known mammalian counterpart. Cdc15, which is essential for completion of mitosis[24], also has no known mammalian homologue; it is related to human MESS1 in the catalytic domain, but has a long dissimilar carboxy-terminal tail.

Diversity of yeast protein kinases

Most of the main vertebrate subfamilies of protein kinases are represented in yeast. For example, in the AGC group there are multiple cAMP-dependent protein kinases (PKAs), a single protein kinase C (PKC), and 70 kDa S6 kinase-related protein kinases, but no protein kinases closely related to cGMP-dependent protein kinase, Rsk or βARK. In the CaMK group, there are Ca^{2+}-calmodulin-regulated and AMP-dependent protein kinases, but no true MLCK, perhaps because myosin-based motility is limited in yeast. In the CMGC group, all the main subfamilies are represented in yeast including CDKs, MAPKs, and GSK3- and Clk-related protein kinases.

New groups. As this is the first analysis of all the protein kinases present in a complete eukaryotic genome, it is appropriate to designate two new groups that are conserved in all eukaryotes, namely the STE11/STE20 group including the MAPKKKs and the STE7 group including the MAPKKs. In addition, several new subfamilies can be established in the 'Other group', including casein kinase I and II, the Ran subfamily and two subfamilies that appear to be unique to budding yeast, Npr/Hal5 and Elm subfamilies. Notably absent, however, are receptor-type protein kinases, and Raf-related protein kinases, which are MAPKKKs that act downstream of receptor protein-tyrosine kinases. Although yeast has several STE11/ MEKK-family MAPKKKs, the absence of Raf-related MAPKKKs and receptor protein kinases in general might reflect the fact that yeast has little need for intercellular communication, other than to respond to mating pheromones, which is

accomplished by G protein-coupled receptors. Many of the protein kinase subfamilies have undergone significant expansion during evolution; for example, there is only one PKC in yeast, but at least nine in vertebrates.

No true protein-tyrosine kinases. As anticipated from many unfruitful sequence-based searches, budding yeast has no members of the true protein-tyrosine kinase family. The progenitor for this family probably arose when multicellular organisms evolved. The driving force behind the evolution of protein-tyrosine kinases was presumably the need for a signaling mechanism for cell–cell communication within a multicellular organism. The concomitant evolution of phosphotyrosine-binding domains that could mediate protein-tyrosine kinase signal-dependent protein–protein interactions must also have been a critical event.

The absence of typical protein-tyrosine kinases, however, does not mean that enzymes of this specificity are completely lacking, and there are several examples of what have been termed dedicated protein-tyrosine kinases. For example, Swe1, a member of the Wee1 family, phosphorylates Tyr19 in Cdc28, negatively regulating the activity of this CDK[25]. Other Wee1 family members can autophosphorylate on serine, threonine and tyrosine, and as a result are commonly known as dual-specificity protein kinases. Although there are no other Swe1-related protein kinases in yeast that could regulate CDK function, there are two protein kinases, Spk1/Rad53/Mec2/Sad1 and Mps1/Rpk1, involved in S and G2 checkpoint control and spindle pole body duplication respectively. These kinases are also dual-specificity protein kinases that, at least *in vitro*, have protein-tyrosine kinase activity[26,27]. Yeast also has true dual-specificity protein kinases in the MAPKK family, which phosphorylate the Thr and Tyr in a TxY motif in the activation loop of members of the MAP kinase family. The existence of three bona fide protein-tyrosine phosphatases in yeast underscores the importance of protein-tyrosine phosphorylation in this organism.

All the protein kinases classified in the protein-tyrosine kinase family based on sequence analysis have experimentally verified tyrosine phosphorylating specificity. However, it is not yet known exactly how tyrosine, rather than serine or threonine, is selected for phosphorylation, despite the availability of several protein-serine and protein-tyrosine kinase

catalytic-domain three-dimensional structures. This means that one cannot exclude the possibility that a novel solitary protein kinase is not a protein-tyrosine kinase until it is tested biochemically. Indeed, mammals have a novel protein-tyrosine kinase, A6, which is totally unrelated in sequence to the conventional protein-tyrosine kinase family[28]. There is an A6-related gene, YGR080W, in yeast, and additional homologues in the human EST database. It will be interesting to test whether these genes encode protein-tyrosine kinases.

Protein kinases found only in yeast

Some subfamilies of yeast protein kinases so far appear to be unique to budding yeast. These include a PKA-related family (YOL100W, YDR490C, YDR466W), a Nek-like family (YNL020, YIL095W, YBR059C) and the Npr/Hal5 family, which has nine members divided into two groups related to Hal5, a protein kinase involved in salt and pH tolerance, and Npr1, a protein kinase involved in regulating transport systems for nitrogen nutrients under conditions of catabolite derepression[29]. If one were to predict what sort of protein kinases might be unique to yeast, those involved in nutrient uptake or resistance to environmental stress would be obvious candidates. The same might be true for the Elm family, where the eponymous protein kinase Elm1 is involved in pseudohyphal growth[30], a process that is unique to yeast. There are also a number of yeast protein kinases that have homologues in other species, but not in vertebrates; for instance the Ran family contains protein kinases closest to the fission yeast Ran1 protein kinase.

There are currently four yeast protein kinases that have no known homologues in other species (Bub1, YKL171W, YGR052W and YPR106W). However, although protein kinases in these families have not been identified in vertebrates so far, several of the yeast protein kinases are close relatives of *Caenorhabditis elegans* protein kinases, where the genome sequence is now nearly 60% complete. Thus, as the human genome sequence progresses, one can anticipate that many of the apparently unique yeast protein kinases will prove to have vertebrate homologues.

Non-conventional protein kinases

In addition to genes in the conventional protein kinase superfamily, there is also a single gene, that encodes a protein in the prokaryotic 'histidine' protein kinase family, Sln1 (Ref. 31). Such proteins autophosphorylate on a histidine residue in response to a specific stimulus, and then this phosphate is transferred to an acceptor signal response protein on an aspartate residue. Given the frequency of such protein kinases in prokaryotes, where more than ten are known in *Escherichia coli*, it is surprising that there are not more protein kinases of this type in yeast. It is interesting, however, that the Sln1 histidine kinase plays a role in the response of yeast cells to osmotic pressure, a response that it is transduced in bacterial cells by the histidine kinase, EnvZ[31]. The YIL042C protein is related to the mitochondrial branched chain α-ketoacid and pyruvate dehydrogenase protein kinases, which are unusual protein-serine kinases that are recognizable members of the 'histidine' protein kinase family[32,33].

There are a number of eukaryotic proteins with bona fide protein kinase activity, such as the *Dictyostelium* myosin heavy chain protein kinase[34], that are structurally unrelated to either the eukaryotic protein kinase superfamily or the prokaryotic signal response protein kinases. Well over 60% of the genes in the yeast genome are of unknown function, and it is certainly possible that some of them will be unconventional protein kinases, such as the A6-related protein discussed above.

The yeast genome has ten genes in the lipid kinase family, which have a domain related to the catalytic domain of the protein kinase superfamily. Some of these are bona fide phosphatidylinositol kinases (e.g. Vps34, Pik1, etc.), whereas others have not been found to have lipid kinase activity. Given that DNA PK, a mammalian protein in this family, is a genuine protein kinase whose activity is stimulated by double-stranded DNA ends, there has been speculation that some of the members of the lipid kinase family are in fact protein kinases[35]. Indeed, it appears likely that the members of this family that are involved in checkpoint function, such as Mec1 and Tel1, are protein kinases. Moreover, another lipid kinase subfamily, which includes the rapamycin-binding proteins, Tor1 and Tor2, is known to autophosphorylate, and it is likely that these are also protein kinases, although their salient substrates remain to be identified.

Perspectives

If we include all the different types of protein kinase encoded by the yeast genome, we reach a total of ~120; a number that is a little lower than the most recent estimate of the number of yeast protein kinases, which was based on the sequencing of chromosome III (Ref. 36). Recent analysis of GenBank and unfinished sequence databases for *C. elegans* (kindly provided by the Sanger Centre, Cambridge, UK) so far reveals 270 unique protein kinases (G. D. Plowman, unpublished). However, although these data represent approximately 60% of the 100 Mb nematode genome, which is about eight times larger than that of yeast, the number of kinases per kb of DNA is only half that predicted from the analysis of the yeast genome. However, the number and length of introns in higher eukaryotes is much greater than in yeast, which decreases the percentage of coding regions.

By extrapolation, we can estimate the number of protein kinase genes in mammals, which have about four times as many genes as *C. elegans*. Based on this estimation, a prediction of more than 1000 protein kinase genes in the human genome still seems a reasonable one[37], particularly if one takes into account the expansion of the protein-tyrosine kinases used for intercellular signalling in higher organisms. By the time the human genome project is completed we will know how accurate this estimate was.

References

1 Garrels, J. I. (1995) *Nucleic Acids Res.* 24, 46–49
2 Hieter, P., Bassett, D. E., Jr and Valle, D. (1996) *Nat. Genet.* 13, 253–255
3 Johnston, M. (1996) *Curr. Biol.* 6, 500–503
4 Waskiewicz, A. J. and Cooper, J. A. (1995) *Curr. Biol.* 7, 798–805
5 Liu, H., Styles, C. A. and Fink, G. R. (1993) *Science* 262, 1741–1744
6 Sia, R. A. and Mitchell, A. P. (1995) *Mol. Cell. Biol.* 15, 5279–5287
7 Krisak, L. et al. (1994) *Genes Dev.* 8, 2151–2161
8 Friesen, H., Lunz, R., Doyle, S. and Segall, J. (1994) *Genes Dev.* 8, 2162–2175
9 Simon, M. N. et al. (1995) *Nature* 376, 702–705
10 Ramer, S. W. and Davis, R. W. (1993) *Proc. Natl. Acad. Sci. U. S. A.* 90, 452–456
11 Leberer, E. et al. (1992) *EMBO J.* 11, 4815–4824
12 Cvrckova, F. et al. (1995) *Genes Dev.* 9, 1817–1830
13 Poon, R. Y. C. and Hunter, T. (1995) *Curr. Biol.* 5, 1243–1247
14 Espinoza, F. H. et al. (1996) *Science* 273, 1714–1717

15 Kaldis, P., Sutton, A. and Solomon, M. J. (1996) Cell 86, 553–564

16 Thuret, J-Y., Valay, J-G., Faye, G. and Mann, C. (1996) Cell 86, 565–576

17 Osmani, S. A. and Ye, X. X. (1996) Biochem. J. 317, 633–641

18 Lu, K. P. and Hunter, T. (1995) Cell 81,413–424

19 O'Connell, M. J., Norbury, C. and Nurse, P. (1994) EMBO J. 13, 4926–4937

20 Kambouris, N. G., Burke, D. J. and Creutz, C. E. (1993) Yeast 9, 141–150

21 Schweitzer, B. and Philippsen, P. (1992) Mol. Gen. Genet. 234, 164–167

22 Kitada, K., Johnson, A. L., Johnston, L. H. and Sugino, A. (1993) Mol. Cell. Biol. 13, 4445–4457

23 Patterson, M., Sciafani, R. A., Fangman, W. L. and Rosamond, J. (1986) Mol. Cell. Biol. 6, 1590–1598

24 Schweitzer, B. and Philippsen, P. (1991) Yeast 7, 265–273

25 Booher, R. N., Deshaies, R. J. and Kirschner, M. W. (1993) EMBC J. 12, 3417–3426

26 Lauze, E. et al. (1995) EMBO J. 14, 1655–1663

27 Zheng, P. et al. (1993) Mol. Cell. Biol. 13, 5829–5842

28 Beeler, J. F. et al. (1994) Mol. Cell. Biol. 14, 982–988

29 Vandenbol, M., Jauniaux, J. C. and Grenson, M. (1990) Mol. Gen. Genet. 222, 393–399

30 Blacketer, M. J. et al. (1993) Mol. Cell. Biol.

13, 5567–5581

31 Posas, F. et al. (1996) Cell 86, 865–875

32 Popov, K. M. et al. (1992) J. Biol. Chem. 267, 13127–13130

33 Popov, K. M. et al. (1993) J. Biol. Chem. 268. 26602–26606

34 Futey, L. M., Medley, Q. G., Cote, G. P. and Egelhoff, T. T. (1995) J. Biol. Chem. 270, 523–529

35 Hunter, T. (1995) Cell 83, 1–4

36 Hunter, T. (1994) Semin. Cell Biol. 5, 367–376

37 Hunter, T. (1987) Cell 50, 823–829

38 Hanks, S. K. and Hunter, T. (1995) FASEB J. 9, 576–596

39 Krogh, A. et al. (1994) J. Mol. Biol. 235, 1501–1531

Protein architecture, dynamics and allostery in tryptophan synthase channeling

Peng Pan, Eilika Woehl and Michael F. Dunn

The $\alpha_2\beta_2$ form of the tryptophan synthase bienzyme complex catalyses the last two steps in the synthesis of L-tryptophan, consecutive processes that depend on the channeling of the common metabolite, indole, between the sites of the α- and β-subunits through a 25 Å-long tunnel. The channeling of indole and the coupling of the activities of the two sites are controlled by allosteric signals derived from covalent transformations at the β-site that switch the enzyme between an open, low-activity state, to which ligands bind, and a closed, high-activity state, which prevents the escape of indole.

THE PHENOMENON OF direct metabolite transfer between sequential enzyme pairs in a metabolic cycle is classified as substrate channeling[1]. The tryptophan synthases from enteric bacteria, with subunit composition $\alpha_2\beta_2$, are the best-characterized examples of substrate-channeling, multienzyme complexes[2,3]. These enzymes catalyse the last two steps (Fig. 1) in the biosynthesis of L-tryptophan (L-Trp). The α-subunit has an $(\alpha/\beta)_8$-barrel folded motif, and catalyses the cleavage of 3-indole-D-glycerol 3'-phosphate (IGP) to indole and D-glyceraldehyde-3'-phosphate (G3P). The β-subunit is a pyridoxal phosphate-requiring enzyme that catalyses the conversion of L-serine (L-Ser) and indole to L-Trp and a water molecule[2]. The

P. Pan, E. Woehl and M. F. Dunn are at the Department of Biochemistry, University of California, Riverside, CA 92521-0129, USA.

β-reaction occurs in two stages; in Stage I, L-Ser reacts with enzyme-bound pyridoxal 5'-phosphate (PLP) to form the quasi-stable α-aminoacrylate intermediate, E(A-A), the species poised for reaction with indole; in Stage II, indole reacts with E(A-A) to form L-Trp (Fig. 1). Efficiency is achieved by channeling the product of the first enzyme (indole) to the second enzyme from the α-site to the β-site through a 25 Å-long tunnel (Fig. 2).

Here, we review recent findings showing that, in the overall catalytic cycle, β-site covalent reactions with substrate trigger allosteric signals that flip the enzyme between open (low activity) and closed (high activity) conformations. These events serve two functions: (1) conversion to the closed state prevents the escape of indole, and (2) switching between activity states couples the catalytic cycles of the two enzymes.

Physical and dynamic constraints are dictated by function

The elegant crystallographic work of Hyde et al.[4] and Rhee et al.[5] on the Salmonella typhimurium tryptophan synthase bienzyme complex ($\alpha_2\beta_2$) has contributed significant advances to our knowledge of structure-function relationships in substrate channeling by multienzyme complexes. Their efforts have provided the first example showing the three-dimensional structure of the molecular machinery required for channeling in a stable multienzyme complex. Out of this structural work and from recent mechanistic studies[6–12], there has emerged the realization that, for efficient substrate channeling to occur between enzyme pairs, a rather stringent set of physical and dynamic constraints must be met. Those evident in the tryptophan synthase example so far are as follows: (1) The architecture of the multienzyme complex must provide a physical structure with dynamic properties that constrain the degrees of freedom of the common metabolite so that transfer from one site to the next is assured. (2) Catalysis at the two active sites must be coupled such that turnover at each site occurs in phase.

The structure[4] of $\alpha_2\beta_2$ partially explains how the first constraint is satisfied; the α- and β-sites of each heterologous dimer are connected by a 25 Å-long tunnel (Fig. 2). The kinetic studies of Dunn et al.[6], Lane and Kirschner[7] and Anderson et al.[8] have established that the tunnel actually functions as the conduit for the transfer of indole between sites.

An interconnecting tunnel is insufficient to ensure channeling

Two additional criteria must be met to achieve efficient phasing of the vectoral transfer of indole with the β-site chemistry (Fig. 1) so that L-Trp is efficiently

Exhibit 15

Review

The protein kinases of *Caenorhabditis elegans*: A model for signal transduction in multicellular organisms

Gregory D. Plowman*, Sucha Sudarsanam*, Jonathan Bingham*, David Whyte*, and Tony Hunter†‡

†The Salk Institute, 10010 North Torrey Pines Road, La Jolla, CA 92037; and *SUGEN, 230 East Grand Avenue, South San Francisco, CA 94080

Caenorhabditis elegans should soon be the first multicellular organism whose complete genomic sequence has been determined. This achievement provides a unique opportunity for a comprehensive assessment of the signal transduction molecules required for the existence of a multicellular animal. Although the worm *C. elegans* may not much resemble humans, the molecules that regulate signal transduction in these two organisms prove to be quite similar. We focus here on the content and diversity of protein kinases present in worms, together with an assessment of other classes of proteins that regulate protein phosphorylation. By systematic analysis of the 19,099 predicted *C. elegans* proteins, and thorough analysis of the finished and unfinished genomic sequences, we have identified 411 full length protein kinases and 21 partial kinase fragments. We also describe 82 additional proteins that are predicted to be structurally similar to conventional protein kinases even though they share minimal primary sequence identity. Finally, the richness of phosphorylation-dependent signaling pathways in worms is further supported with the identification of 185 protein phosphatases and 128 phosphoprotein-binding domains (SH2, PTB, STYX, SBF, 14-3-3, FHA, and WW) in the worm genome.

Reversible protein phosphorylation plays a central role in regulating basic functions of all eukaryotes such as DNA replication, cell cycle control, gene transcription, protein translation, and energy metabolism. Protein phosphorylation is also required for more advanced functions in higher eukaryotes such as cell, organ, and limb differentiation, cell survival, synaptic transmission, cell–substratum and cell–cell communication, and to mediate complex interactions with the external environment. Because aberrant protein phosphorylation is commonly the cause of cancer and other human diseases, a comprehensive knowledge of the key enzymes that regulate these functions can provide the basis for novel therapeutic intervention strategies.

The genomic revolution promises to provide a new paradigm for drug discovery, allowing one to selectively target the molecular basis of human disease. The completion of the *Caenorhabditis elegans* genome sequence gives us an opportunity to decipher the molecular nature of its signal transduction machinery. Several global analyses of proteins and protein domains present in *C. elegans* have been presented elsewhere (1–4), revealing that protein kinases comprise the second largest family of protein domains in worms. The three most frequently occurring protein domains found in worms are seven transmembrane chemoreceptors (650 domains, 3.5% of genome), protein kinases (496 domains, 2.6% of genome), and zinc finger C4 domains, including nuclear hormone receptors (275 domains, 1.4% of genome). A more in-depth analysis has been performed on the 535 worm proteins containing zinc-binding

domains, including the C4, C2H2, and C3HC4 ring finger types (3), and on the 83 worm homeobox transcription factors (4). Here, we present a comparative analysis of the enzymes and adaptor molecules that are the key components of the protein phosphorylation signaling network present in *C. elegans*.

Identification and Classification of *C. elegans* Protein Kinases. To identify worm protein kinases, we first used an HMMER 2.1.1 (http://hmmer.wustl.edu/) profile search against the 19,099 predicted worm proteins, the finished and unfinished *C. elegans* genomic sequence, and the worm chromosome assemblies. The nucleic acid databases were first translated in all six frames, and ORFs longer than 30 amino acids were parsed into a relational database. We generated a hidden Markov model based on 70 representative yeast and human protein kinases whose catalytic domains share <50% sequence identity with each other (5). Using a similar strategy, additional profiles were generated for other protein kinase-like domains (phosphoinositide kinases, atypical A6 kinases, diacylglycerol kinases, aminoglycoside resistance kinases, and microbial kinases), protein phosphatases, and domains capable of specifically binding to phosphotyrosine (P.Tyr) or phosphoserine/threonine residues (SH2, PTB, STYX, SBF, 14-3-3, FHA, and WW domains). Scripts were written for reassembly of contiguous exons identified from genomic sequence to generate the predicted catalytic domain sequence of each kinase. Pairwise BLAST 2.0 (ftp://ncbi.nlm.nih.gov/blast/executables/) analysis was performed to identify redundant entries, and putative protein kinases with low profile scores were manually inspected to determine whether they should be included in subsequent analyses.

This analysis generated a nonredundant list of 493 protein kinase-like proteins and 21 protein kinase gene fragments from worms. This number will continue to increase as the genome is completed and the final assembly of the six worm chromosomes is achieved. Of note, we found >40 kinase domains from genomic analysis that were absent in the 19,099 worm protein dataset. These omissions result from the limitations of current protein prediction algorithms. Furthermore, numerous entries had apparent internal deletions of conserved kinase motifs, likely attributable to inappropriately assigned splice junctions. These sequences were corrected before further classification. Many of the 19,099 proteins were alternate isoforms of the same gene, in which case we included
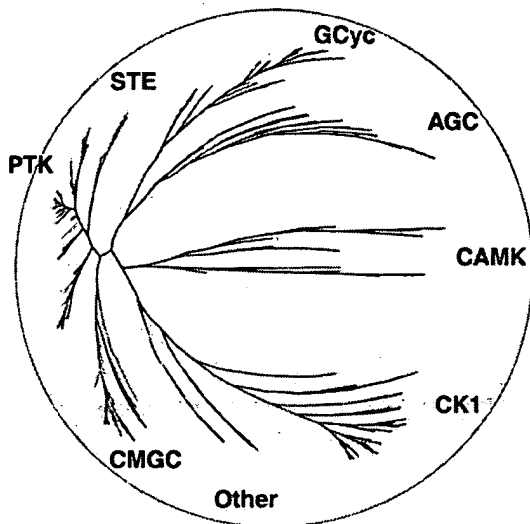

Fig. 1. Hyperbolic tree representation of *C. elegans* protein kinases. Major protein kinase groups are labeled in different colors. A JAVA tool for viewing this dendrogram can be found at www.kinase.com.

Table 1. Summary and classification of phosphoprotein signaling molecules in worms, budding yeast, and humans

Superfamily	Group	Worm	Fragments worm	Yeast	Human
Protein kinase	AGC	30	1	17	100
	CAMK	32	0	21	83
	CKI	87	7	4	5
	CMGC	42	0	24	62
	Other	62	6	29	163
	STE	28	0	15	63
	PTK	92	5	0	100
	"Worm"	27	2	0	0
	"Yeast"	0	0	4	0
	"Microbial"	7	0	6	5
	Atypical	4	0	4	11
	All	411	21	124	592
PK-like	Gcyc	26	0	1	8
	PIK	12	0	10	20
	DAGK	7	0	2	8
	YLK1	30	0	0	0
	Choline K	7	0	2	2
	All	82	0	15	38
Phosphatase	cPTP	57	4	3	25
	RPTP	26	14	0	22
	DSP	26	0	16	51
	STP	65	0	18	21
	IPP	11	0	7	7
	All	185	18	44	126
PP-Binding	SH2	73	1	1	>137
	PTB	16	0	0	>47
	STYX (DSP)	1	0	5	2
	SBF (MTM)	2	0	0	3
	14-3-3	3	0	2	>6
	FHA	11	0	14	>20
	WW	22	0	5	>32
	All	128	1	27	>247
Other	Cyclins	34	0	23	>21

only one of the proteins in our final assessment. In determining the total number of protein kinases, the three proteins determined to contain dual catalytic domains were only counted once. Many of the protein ORFs truncated the extremities of the kinase domain proteins, frequently because of their location near the end of a cosmid clone. In these cases, we searched for N- or C-terminal domains on adjacent cosmids to assist in the subsequent classification. One challenge of genomic data mining is the presence of sequence repeats. Tandem repeats and inverted repeats account for 2.7 and 3.6% of the worm genome, respectively. In addition, worms contain large regions of tandem gene duplication, ranging from hundreds of bases to >100,000 bases (1). In some cases, the genes encoded within these regions are duplicated and have nearly identical sequences. Therefore, until the chromosome sequences are fully assembled, data-mining approaches may exclude some of these duplicated genes.

A multiple sequence alignment was generated from the predicted catalytic domains of 398 of these protein kinase, which share >15% amino acid identity with other entries. The aligned proteins were then clustered by using parsimony analysis, and the results were displayed as rooted and unrooted cluster dendrograms, and as kinase "retinograms" or hyperbolic trees using a JAVA display tool (Fig. 1 and www.kinase.com). The protein kinases were then classified into several kinase groups and families, based on relatedness within the kinase catalytic domain to other worm, yeast, and vertebrate protein kinases. Further classification was performed by searching for noncatalytic domains linked to the kinase domain, including predicted transmembrane regions, SH2 domains and SH3 domains, and Ig and fibronectin Type III domains.

Table 1 presents a summary of our classification of the 411 protein kinases and 82 protein kinase-like motifs. A more detailed table of these proteins, along with basic informatics tools for retrieval and alignment of these sequences can be found on our web site at www.kinase.com. Table 1 also summarizes the results of a similar analysis of the completed yeast genome and of an ongoing effort from publicly available human expressed sequence tag and genomic databases. From this classification, we can now determine which protein kinases are conserved between yeast and worms, we can speculate on the origin of the protein kinase superfamily, and we can identify kinases that are yeast-specific and those that are restricted to higher eukaryotes. We tentatively identify "worm-specific" protein kinases, based on their absence from current

mammalian expressed sequence tag and nucleic acid databases. However, a final assessment will have to await completion of the *Drosophila* and human genome sequences. We also elaborate on some of the protein kinases and signaling pathways that evolutionarily appear only in more complex organisms such as vertebrates.

In this review, we use the term "orthologues" to refer to proteins of different species that are believed to have a common ancestor and have an evolutionarily conserved function. Orthologous proteins typically have similar domain structure and share extended sequence similarity outside of their catalytic domains. Homologous proteins also share extended sequence similarity, but to a lesser degree than orthologues, and are not expected to complement one another functionally. However, within large protein superfamilies such as protein kinases, G protein coupled receptors, and nuclear hormone receptors, there is not a single expectation value that can be used to categorize all members definitively, and final classification will require experimental validation.

Yeast- and Fungal-Specific Kinases. The first complete eukaryote sequence, that of the budding yeast *Saccharomyces cerevisiae*, was reported in 1996 (6). Shortly thereafter, we presented a comprehensive analysis and classification of yeast protein kinases (7). Now, with the availability of a second eukaryotic genome, *C. elegans*, we can perform a similar analysis and make more informed general-

izations on which of these protein kinases are unique to yeast or fungi, and also on which protein kinases evolved during the emergence of multicellular organisms and are therefore not represented in yeast or fungi.

We now identify a total of 24 yeast-specific protein kinases and an additional 3 that are currently restricted to yeast and worms. Originally we defined four protein kinase subfamilies, containing a total of 18 members, to be yeast specific [protein kinase A (PKA)-related, RAN, ELM, and NPR/HAL5 families]. These remain yeast- or fungal-specific, as no close homologues are present in worms, and none have yet been described in vertebrates. However, the ELM family could be considered as a subfamily of the CAMK group. Rim15 is a yeast-specific kinase that is related to *Schizosaccharomyces pombe* Cek1, and its similarity to budding yeast YNL161w places it as a distant member of the NDR family kinases. Two other protein kinase subfamilies, containing a total of five members, were originally recognized as having only distant homologues in higher organisms (NEK-like and PIM-like families). The prototype of the NEK-like family, YNL020C, has a homologue in worms, but not in mammals, although its C-terminal tail has a predicted coiled-coil structure related to numerous mammalian protein kinases (e.g., SLK/PLKK, TAK1). The two yeast PIM-like family members have catalytic domains related to worm and mammalian protein kinases, but have a unique N-terminal domain.

Members of the NPR/HAL5 family are involved in ion homeostasis, polyamine transport, nutrient uptake, and response to nitrogen starvation, whereas Elm1 initiates a protein kinase cascade controlling pseudohyphal growth (8). Members of the RAN family are related to fission yeast Ran1/Pat1, which regulates the switch between vegetative growth and meiosis. Because these are fungal-specific responses, it is not surprising that these protein kinases are restricted to lower eukaryotes.

A second set of "unique" yeast protein kinases was originally defined because they had no close homologues in other species (7). Most of these yeast protein kinases now have both worm and vertebrate orthologues (Cdc5, Ipl1, Ire1, Vps15, YGL180W/Apg1, Swe1, Spk1, Gcn2, YBR274W, YGR262C, and Bub1). Exceptions among this list of unique yeast protein kinases are YPL236C and Mps1, which have orthologues in humans, but not in worms; YKL116C, which is distantly related to the EMK-family, yet has only weak homologues in worms and humans; and YKL171W, YGR052W, and YPR106W, which remain yeast specific protein kinases. Two sequences that were excluded from our previous analysis of yeast protein kinases deserve mention. The budding yeast protein Iks1 can be classified as a yeast-specific protein kinase because it still has no homologues in worms or other species whereas another yeast kinase-like sequence, SCY1, has orthologues in *C. elegans* and *Arabidopsis*, but none thus far in vertebrates. A *S. pombe* protein, which is distantly related to SCY1, also has a single worm orthologue.

Worm-Specific Protein Kinases. Which protein kinases are specific to worms? Protein kinases that are absent from yeast yet present in worms are likely to be involved in the complex signal transduction pathways that are required for the existence of multicellular organisms. These might include protein kinases involved in cell–substratum and cell–cell adhesion, transmembrane signaling in response to humoral factors, protein kinases involved in cell survival or programmed cell death, and protein kinases whose signals regulate metazoan-specific transcription factors, particularly those containing Zn-finger domains.

In the absence of complete genome sequences of other multicellular eukaryotes, we tentatively classify 165 protein kinases (plus 9 protein kinase fragments) as worm-specific. The majority (134, 80%) fall into three groups (CK1, FER, and KIN-15) whereas the others are distant members of common protein kinase families or belong to worm specific subfamilies. Five protein kinase subfamilies, containing a total of 12 members, can tentatively be defined as worm-specific (C04G2.10, K08B4.5, K09C6.7, R107.4, and ZK177.2-families). An additional 15 unique worm protein kinases are also identified, which to date have no close homologues in yeast, worms, or in higher organisms. However, mammalian homologues of some of these worm protein kinases are already beginning to appear in publicly available expressed sequence tag databases, and assignment of a protein kinase as being truly worm-specific will have to await the completion of the *Drosophila* and human genome sequences.

Members of four other protein kinase or kinase-like subfamilies are disproportionately represented in worms compared with humans. Clusters of 5–9 members of each of these families are localized to short regions (<1 megabase) of chromosomes II and IV, suggesting they may each have expanded as a result of extensive tandem gene duplication. The chromosomal density of protein kinases is graphically depicted on our web site at www.kinase.com. The four gene families are the CK1-family, the KIN-15-family of receptor protein-tyrosine kinases, the FER-family of cytoplasmic protein-tyrosine kinases, and the kinase-like domains of the receptor guanylyl cyclases.

CK1 family. The worm genome contains 87 CK1 (casein kinase I) members (plus 7 additional partial catalytic domains) whereas there are only 4 known members in budding yeast and 6 in humans. Genetic evidence from the yeast homologues suggests CK1s may be involved in DNA repair and cell division, and mammalian CK1s have been shown to phosphorylate p53 in G1 and G2, possibly affecting cell sensitivity to DNA damage at these checkpoints (9). Little is known regarding the function of CK1s in worms, but the enormous arborization and diversification of this kinase family may be an adaptation allowing for enhanced DNA repair in response to excessive exposure to environmental mutagens.

KIN-15/16 family. C. elegans contains 16 members of a unique family of receptor protein-tyrosine kinases whose presence to date is restricted to this species. These transmembrane proteins have unusually short (<50-aa) extracellular domains, and many are clustered within the genome, as though they arose through tandem gene duplication. The prototype members of this family, KIN-15 and KIN-16, are expressed in the hypodermal syncytium, which expands by cell fusion during larval development (10). Compared with wild-type worms, KIN-15 and KIN-16 deletion mutants produce fewer embryos and rarely develop into adults, but, when they do mature, they typically exhibit extrusion of the gonads through the vulva (11). Therefore, KIN-15/16 appear to be essential genes, yet may undergo variable compensation by 1 of the 14 other homologues. One of the KIN-15 clusters is interspersed with chitinase genes, which are known to function in cell wall morphogenesis during the molting process and in fungal resistance. Expansion of this region may have been necessary during evolution to facilitate this aspect of larval development. An alternative function for KIN-15-family kinases is suggested by the fact that overexpression TKR-1 (C08H9.5) causes a 40–100% extension of life expectancy in worms (12). Unlike other life extension (*age*) mutants, TKR-1 transgenics do not form dauers, and their longevity has been attributed to an increased resistance to ultraviolet and thermal stress.

FER family. The worm genome contains 42 members (plus 2 additional partial catalytic domains) of the FER-family of single SH2-containing cytoplasmic protein-tyrosine kinases. Most of these genes are interspersed throughout the worm genome; however, nine members reside within a 1.1-megabase region on chromosome IV. Unfortunately, no literature is available on the function of any of these protein kinases in worms, but the two mammalian homologues, FER and FES, have been demonstrated to play a role in cell adhesion, to signal downstream of cytokine receptors, and to function as oncogenes (13). Conceivably, additional human representatives will be revealed on completion of the human genome sequence, possibly with restricted expression. Alternatively, their function may be replaced in humans by expansion and

REVIEW

diversification of non-FER cytoplasmic protein-tyrosine kinases, of which worms have only 10 whereas humans have at least 34. Most evident is a dramatic expansion of SRC-family kinases and emergence of ZAP70 and JAK family kinases in higher eukaryotes that are not found in the worm genome.

Conserved Metazoan Protein Kinase Signaling Transduction Pathways. Worms provide an elegant model system for studying signal transduction. This transparent animal is comprised of 959 somatic cells plus 131 cells destined for programmed cell death. The *C. elegans* hermaphrodite contains 302 neurons and 81 muscle cells and has a brain, reproductive system, and digestive tract (ref. 14; http://dauerdigs.biosci.missouri.edu/Dauer-World/Wormintro.html). It provides a complex yet tractable system for studying development, metabolism, aging, and behavioral responses to a number of stimuli. Regulation of many of these processes is carried out through signal transduction pathways that are also present in humans. Not surprisingly, all of the major protein kinase groups found in worms are also conserved in humans (15). The number of protein kinases classified into each major group from yeast and worms, along with a current estimate from humans, is provided in Table 1. These numbers represent a current analysis, but new protein kinases are being discovered every month as the worm genome sequencing project continues. Some of these entries may also represent pseudogenes containing frameshifts that result in incomplete translation into a full kinase catalytic domain.

AGC Group. The AGC group of worm protein kinases contains representatives of many of the known types of cyclic nucleotide-dependent, NDR or DBF2, and ribosomal S6 kinase families. Worms also contain members of the cGMP-dependent kinase (PKG), RSK, and G-protein coupled receptor kinase families that are absent from budding yeast. Two of the S6 kinase members have dual catalytic domains similar to vertebrate RSK enzymes, where the N-terminal domain clusters into the AGC group and the C-terminal kinase domain is most related to the CaMK group. Worms have four members of the AKT family, two being close orthologues of mammalian AKT1/PKB/RACα, and two related to the AKT upstream kinase, PDK1. AKT is a mammalian protooncoprotein regulated by phosphatidylinositol 3-kinase (PI3-K), which appears to function as a cell survival signal to protect cells from apoptosis (16). Insulin receptor, RAS, PI3-K, and PDK1 all act as upstream activators of AKT whereas the lipid phosphatase PTEN functions as a negative regulator of the PI3-K/AKT pathway (17). Downstream targets for AKT-mediated cell survival include the proapoptotic factors BAD and Caspase9 and transcription factors in the forkhead family, such as DAF-16 in the worm. AKT is also an essential mediator in insulin signaling, in part because of its use of GSK-3 as another downstream target. Each of these components of the AKT/PI3-K pathway is conserved in worms, providing a powerful system for genetic dissection of a major cell survival signal.

The cAMP-dependent protein kinases (PKA) consist of heterotetramers comprised of two catalytic (C) and two regulatory (R) subunits, in which the R subunits bind to the second messenger cAMP, leading to dissociation of the active C subunits from the complex. Worms have two PKA catalytic domains and two regulatory subunit genes (R07E4.6 and ZK370.4). Additional cNMP-binding domains are present in the two worm representatives of the PKG family, in several cNMP-gated ion channels, and in a cAMP-regulated guanine nucleotide exchange factor (T20G5.5).

CaMK Group. In the CaMK group, the most abundant representatives include Ca^{2+}/calmodulin-regulated and AMP-dependent protein kinases and EMK-related kinases. Worms also contain members of the death-associated protein kinase, mitogen-activated protein kinase (MAPK)-associated protein kinase, myosin light chain kinase, and phosphorylase kinase families that are absent

from budding yeast. All of these protein kinase families have likely evolved as a result of the demands of multicellularity and the emergence of complex organ systems. For example, even though yeast have myosin homologues, they lack myosin light chain kinases. These protein kinases have presumably evolved to regulate myosin during muscle contraction. A worm contig still under construction appears to contain a phosphorylase kinase catalytic γ subunit orthologue, consistent with the presence of two orthologues of the noncatalytic phosphorylase kinase α subunits, which facilitate calmodulin-binding and are required for activation of the mammalian holoenzyme.

Worms lack a homologue of the mammalian Trio-family kinases. Trio is a large multidomain protein kinase containing Ras and Rho guanine exchange factor domains in addition to PH, SH3, and spectrin domains (18). Trio may link Rho and Rac signaling pathways and appears to be involved in the cytoskeletal changes required for cell migration. Although worms lack a member of this kinase family, they do have at least two proteins related to the entire noncatalytic domain of Trio (UNC-73 and F55C7.7).

We have also identified a forkhead homology (FHA) domain-containing CHK2 orthologue in worms. In yeast, Spk1/Rad3 functions as a DNA damage checkpoint sensor through its FHA domain interacting with phosphorylated Rad9 (19). Although no close orthologue of Spk1 exists in metazoans, this function appears to be replaced by CHK2/CDS1, which is phosphorylated in response to DNA damage and may work in conjunction with CHK1 kinase to phosphorylate CDC25C to prevent premature entry into mitosis (20).

CMGC Group. In the CMGC group of serine/threonine kinases, all of the main subfamilies are conserved between yeast, worms, and mammals, including cyclin-dependent kinase (CDK), MAPK, GSK-3, and CLK. An exception is the RCK family, which is absent from yeast but has two members in worms and at least seven in humans. The worm RCK kinases are most similar to mammalian MAK, or male germ cell-associated kinase, which has been implicated in spermatogenic meiosis and in signal transduction pathways for sight and smell. Worms have 14 CDKs (compared with 5 CDKs in yeast) including orthologues of CDC2, CDK3, CDK5, CDK7, and CDK8, and contain 34 cyclins, compared with 23 in budding yeast (Table 1), including one cyclin H orthologue, which we predict will interact with worm CDK-7 to generate a functional cyclin-activated kinase.

Worms have 14 MAPKs, compared with 6 in yeast and at least 14 in humans. The worm MAPKs include representatives of each of the major types of MAPKs: ERK/MAPK, JNK/SAPK1, p38/SAPK2, BMK/ERK5, and NEMO-like kinase (NLK) (21). In budding yeast, three protein kinase families (the prototypes being Ste20, Ste11, and Ste7) function upstream of the MAPKs to generate at least four distinct MAPK signaling pathways that mediate the response to pheromone, nutritional starvation, and cellular or osmotic stress. In multicellular organisms, these MAPK cascades have evolved to mediate responses to diverse signals including growth factors, mitogens, hormones, and cytokines, in addition to the more primitive stress responses to anoxia, heat shock, and osmotic stress.

STE Group: MAPK Pathways. The STE family refers to the three classes of protein kinases that lie sequentially upstream of the MAPKs. In worms, this group includes 10 STE7 (MEK or MAPKK) kinases, 2 STE11 (MEKK or MAPKKK) kinases, and 12 STE20 (MEKKK) kinases. Based on the number of MAPK and STE-family kinases in *C. elegans*, we predict worms will contain at least 8–10 MAPK pathways. In humans, several protein kinase families that bear only distant homology with the STE11 family also operate at the level of MAPKKKs, including RAF, MLK, TAK1, and COT. Except for COT, worms also have orthologues of each of these kinases. Because crosstalk takes place between protein
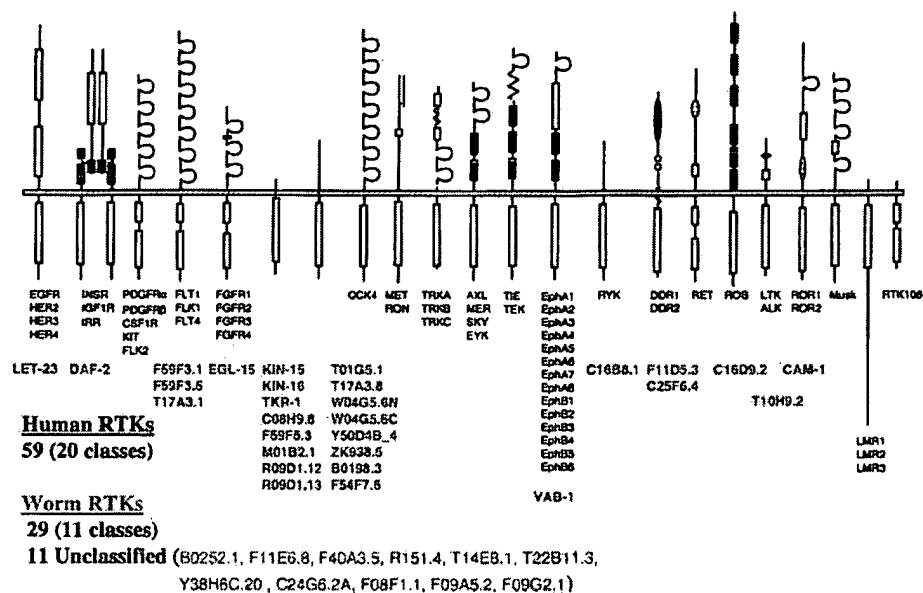
Fig. 2. Schematic representation of the human and C. elegans receptor protein-tyrosine kinase families. Catalytic domains are shown in yellow. The names of the human RTKs are in black, and the names of the worm RTKs are in red.

kinases functioning at different levels of the MAPK cascade, the large number of STE family kinases could translate into an enormous potential for upstream signal specificity and diversity.

Protein-Tyrosine Kinase Group: Receptor Protein-Tyrosine Kinases (RTKs).

The largest group of protein kinases in worms are the protein-tyrosine kinases (PTKs), with 92 members and 5 fragments. We predict this will also remain the largest group of protein kinases in higher eukaryotes, including humans, where the current count is ≈100. These numbers are impressive when one considers that this family is absent from budding yeast. Yeast, however, do have a "budding" tyrosine phosphorylation signaling system, with several dual-specificity kinases (CLK-like, Ste7/MEK family, Swe1, Spk1/Rad53, Mps1), an atypical A6 PTK, 3 protein-tyrosine phosphatases, 16 dual-specificity and low molecular weight phosphatases, and 6 "infant" P.Tyr-binding proteins comprising an apparently nonfunctional SH2 domain protein and 5 phosphatase-like STYX domains. Budding yeast lack PTB domains, and none of the six potential P.Tyr-binding domains have been functionally verified.

The 92 worm PTKs can be further classified into receptor protein-tyrosine kinases (RTKs) and cytoplasmic protein-tyrosine kinases (CTKs) based on the presence or absence of a transmembrane domain and SH2 or SH3 domains. Based on this analysis, the worm genome contains 40 RTKs and 52 CTKs. The 40 RTKs include 16 members of the worm-specific KIN-15-family, 13 RTKs with orthologues representing 10 of the 20 families of human RTKs, and 11 RTKs that remain unclassified with no identifiable mammalian counterpart (Fig. 2). Genetic studies in worms support the classification of five of these worm–human pairs, including LET-23/EGF receptor, DAF-2/insulin receptor, EGL-15/FGF receptor, CAM-1/ROR1 receptor, and VAB-1/EPH receptor, and each of these orthologous pairs mediates similar functions in worms and man, with specificity for epidermal, metabolic, mesodermal, and neuronal signaling pathways.

Based on extracellular domain homologies, we also predict three worm orthologues of PDGFR/FLK/VEGFR, two for DDR, and one each of RYK, ROS, and LTK/ALK. Two of the unclassified RTKs have weak similarity to MET, but not enough to warrant inclusion into this family. Missing in C. elegans are TRK/nerve growth factor receptors, AXL/TYRO3, TIE/angiopoietin recep-

tor, RET/GDNF receptor, and MUSK family members. Identification of three members of the PDGFR/VEGFR family is significant, as they emerged only through analysis of the genomic data and failed to be properly identified from a recent analysis of the predicted 19,099 proteins. Each of these receptors contains multiple Ig-like extracellular domains and a single split kinase domain with closest homology to human FLT1/VEGFR1 and the C. elegans KIN-15 family. However, they are likely to represent early ancestors to both the FLK and PDGFR kinase lineages. Expression of the mammalian FLK/VEGFR RTKs is primarily restricted to endothelial cells, and they play important roles in the early differentiation of hematopoietic and endothelial lineages as well as in normal and pathologic angiogenesis in the adult. However, because worms lack a vasculature, the function of these receptors is not obvious. The formation of mammalian vasculature is reminiscent of the process by which networks of branching tubes develop into the lung and kidneys. Invertebrate VEGFRs may therefore be involved in processes that later evolved into a program for limb and organ development in vertebrates.

Surprisingly little is known about how the ligand-activated VEGFRs mediate these effects. Gene knockout studies in mice suggest that A-RAF or MEKK1 may function downstream of VEGFRs, and recent evidence implicates the involvement of STATs (signal transducer and activator of transcription) in VEGFR signaling (22). Genomic analysis reveals two worm orthologues of STATs (Y51H4, Y43D4 unfinished and F58E6.1), making the VEGFR-STAT association an attractive area for further investigation. STATs contain an SH2 domain, a tyrosine phosphorylation domain, and a DNA-binding domain, and function in a unique JAK-STAT signaling pathway. Extensive studies in mammalian systems have established a model in which JAK kinases are constitutively bound to the cytoplasmic portion of cytokine receptors and are activated on receptor dimerization, facilitating recruitment of STATs to the receptor complex. Subsequent STAT phosphorylation leads to their dimerization and translocation to the nucleus, where they function directly as transcription factors. Drosophila and Dictyostelium STATs both regulate cell division and pattern formation (23, 24). Drosophila STAT has been genetically and biochemically linked to a JAK-STAT signal transduction pathway that regulates pair-rule genes and hematopoiesis. Dictyo-

stelium STAT plays an essential role during the differentiation and aggregation of independent spore cells into stalk cells in response to the chemical signal referred to as differentiation-inducing factor. Furthermore, the *Dictyostelium* AX2 PTK has a second kinase-like domain found only in JAK-family kinases, suggesting the existence of a signaling network similar to that in flies and mammals. However, worms have no cytokines, no cytokine receptors, and no JAK-family kinases. Possibly, the JAK kinase function is replaced by a worm-specific FER kinase, or the STATs may have initially evolved to serve an alternative purpose. Mammalian STATs are also involved in signaling through receptors for growth factors such as EGF, PDGF, VEGF, and angiopoietins. Because the EGF and VEGF signaling systems are present in worms, it is tempting to speculate that these represent the primordial raison d'être for STATs.

In general, related RTKs bind related ligands. In humans, there are at least 12 ligands, encoded by 10 genes, that have been shown to bind selectively to at least one of the four known EGFR-family members. Each of these ligands shares a conserved six-cysteine pattern in its receptor binding domain. In worms, LIN-3 has been shown to function as a LET-23 ligand. Although EGF motifs are prevalent in worms, we have identified three EGF-like proteins (F58G4.4, Y69H2.2, YG70G10A.2) that, in addition to the six cysteines, conserve many of the crucial receptor-binding residues and are juxtaposed next to a putative transmembrane domain, in a pattern similar to all known EGFR-family ligands. Worms also contain at least 3 FGF-like ligands, 12 insulin-like ligands (many more on inclusion of relaxin-related ligands), 2 distant homologues of VEGF, and 4 ephrin-related ligands, some of which would be predicted to bind to their cognate receptors.

Orthologues of other RTK ligands prove more difficult to identify empirically. We see no evidence for a bona fide PDGF or NGF, and searches for ligands for MET, TIE, and AXL-family RTKs are confounded by their similarity to plasminogen, fibrinogen, and fibrillin, respectively. Furthermore, except for weak homologues of MET, these three RTK families are absent from worms. Nevertheless, the significance of a putative Ang2-like protein (Y43C5A.2) in the absence of a TIE-family RTK remains to be determined.

Protein-Tyrosine Kinase Group: CTKs. Most of the 52 CTKs in worms belong to the single SH2-containing FER family. Of the remaining 10 CTKs, there are 2 orthologues of the SH3-containing ACK, and 1 each of FYN (SRC family CTK), FRK, CSK, ABL, and FAK, plus 3 unclassified CTKs. In vertebrates, CSK negatively regulates FYN-family kinases by phosphorylation of a C-terminal tyrosine facilitating a conformational change through an intramolecular SH2-P.Tyr interaction (25). We predict a similar functional interaction between worm FYN and CSK. Co-evolution of this regulatory pair suggests even early metazoans required a means to dampen signaling through CTKs. Notably absent in worms are protein kinases related to the ZAP70 and JAK CTKs, whose primary role in mammals is in signaling through the T cell and cytokine receptors, both of which represent more specialized pathways not present in worms. Humans have eight SRC-family kinases whereas worms have only one. This redundancy has confounded efforts to dissect out the precise role of these CTKs in human biology, often requiring "triple knockouts" to demonstrate a deficiency. The simplicity of non-FER-like CTKs in worms may be helpful in placing these CTKs within specific signaling cascades.

Protein-Tyrosine Kinases: Adaptor and Docking Molecules. Ligand activation of RTKs results in tyrosine phosphorylation of both the receptor itself (autophosphorylation) and of downstream substrates. These phosphorylated tyrosines then function as attachment sites for proteins containing SH2 and other P.Tyr-binding domains. We have identified 74 proteins containing a total of 77 SH2 domains in worms. The majority of these SH2 domains are in CTKs, two are present in a SHP2-related PTP, and the remainder

are predicted to represent orthologues of a variety of adaptor molecules, including phospholipase Cγ, CBL, CIS4/SOCS5, CRK, NCK, SEM-5/GRB2, SHC, tensin, STAT, and VAV. Worms also contain at least 16 PTB domains, which in some cases have been found to interact specifically with tyrosine phosphorylated proteins. Worm PTB-containing proteins include orthologues of SHC, which also contains an SH2 domain, neuronal transmembrane protein X11, and an insulin receptor substrate (IRS) family member. The mammalian X11 PTB domain does not to bind to P.Tyr, so we anticipate only a few of these worm domains will function as P.Tyr-binding domains. Additional potential phosphoprotein-binding domains identified in worms include three 14-3-3 domains, 22 WW domains, and 11 FHA domains.

IRS-1 and IRS-2 are major substrates of the insulin receptor RTK in mammals, and disruption of IRS-2 in mice leads to metabolic defects similar to diabetes. Worms have multiple insulin-like peptides, a receptor, and an IRS orthologue, demonstrating the early origins of metabolic regulation in multicellular organisms. The presence of such a diverse array of adaptor molecules underscores the utility of worms as a model for understanding mammalian signal transduction.

Other Protein Kinases. Approximately 15% of the worm protein kinases do not fall into one of the six major groups but include smaller families with representatives in higher eukaryotes, including CHK1, DYRK, MLK, TAK1, PIM, RAF, STKR, and the mitotic kinases (BUB1, AURORA, PLK, and NIMA/NEK). Recent genetic and biochemical data place TAK1 (transforming growth factor β-associated kinase) on a MAPK-like pathway at the level of a MAPKKK acting upstream of the MAPK-family member NLK. The worm orthologues of TAK1 and NLK regulate Wnt-mediated cell polarization during embryogenesis (21). Biochemical data also demonstrate that this MAPK-like pathway negatively regulates Wnt signaling because NLK phosphorylates the TCF/LEF HMG transcription factors, thereby inhibiting Wnt-regulated binding of the β-catenin-TCF complex to DNA. Both of these pathways are conserved between mammals and worms. The likely orthologous human/worm pairs on the TAK1 MAPK-like pathway include TAK1/MOM-4, NLK/LIT-1, and TCF4/POP-1. Upstream regulators may include TGFβ1/DBL-1, TGFβ type I receptor/SMA-6, TGFβ type II receptor/DAF-4 (worms have three receptor serine kinases). Additional components of the Wnt-signaling pathway, such as cadherin, the adenomatous polyposis coli tumor suppressor gene (APC), disheveled, and GSK-3 kinase are also present in worms, suggesting that there may be a primordial connection between polarized control of cell division/migration and cellular transformation in vertebrates (26).

Microbial-Like Kinases: Origin of Protein Kinases? The availability of the sequence of the first complete metazoan genome, combined with the sequence of budding yeast and several prokaryotic and *Archaea* genomes, provides an excellent opportunity to reassess current theories on the evolutionary origin of protein kinases. Pkn1 is a bacterial protein kinase-like sequence first described in the Gram-negative bacteria *Myxococcus xanthus*, which functions in growth and differentiation and in the ability of this prokaryote to form a fruiting body in response to nutrient starvation. Pkn-related proteins are present in other prokaryotes, including *Streptomyces*, *Bacillus*, *Mycobacterium*, *Pseudomonas*, *Chlamydia*, and *Synechocystis*, where they are involved in virulence, secondary metabolism, sporulation, and complex growth cycles (27). However, there are no Pkn homologues in bacteria with less complex life cycles, such as *Escherichia coli*, and *Haemophilus influenzae*, or in any *Archaea*, suggesting they may have been acquired by horizontal transmission from an early eukaryote, and are unlikely to represent the ancestral founders to protein kinases.

In our kinase profile searches of the worm genome, we detected several entries with low profile scores, yet with significant (E value $< 10^{-2}$) random expectation (E) values. Most of these contained similarity to kinase subdomains I, II, and VI, containing

the consensus GxGxxGxV, VAVK, and HxDxxxxN motifs, respectively. Upon further analysis, many of these entries could be classified into distinct families designated ABC1, RI01, YGR262, diacylglycerol kinase, choline/ethanolamine kinases, and the YLK1-antibiotic resistance kinases. The first three families are named after their prototypic members in S. cerevisiae (27).

Worms contain three proteins related to the budding yeast ABC1. The yeast protein is required for the assembly of the mitochondrial cytochrome c reductase complex, which functions as an electron carrier during oxidative phosphorylation to generate ATP (28). ABC1 homologues are present in numerous prokaryotes, including Mycobacterium, Clostridium, Rickettsia, Synechocystis, Azobacter, and Enterobacteriaceae such as E. coli and Providencia stuartii, in addition to the Archaea, Methanobacterium. ABC1-like proteins are also present in eukaryotes, including fission yeast, Arabidopsis, worms, and mammals. Although ABC1 homologues are absent from bacteria such as Mycoplasma, Bacillus, Haemophilus, Helicobacter, and spirochetes, their presence in other prokaryotes, Archaea, and eukaryotes positions them as likely representatives of the primordial protein kinase, which was the progenitor of the eukaryotic protein kinase family. Based on their recognized role in mitochondrial ATP production and because they maintain many of the structurally important residues and motifs involved in ATP binding, the ABC1-family proteins may either bind ATP or act as phosphotransferases. Conceivably, the ABC1 proteins transfer phosphate to proteins as part of a feedback loop to sense mitochondrial ATP levels.

The RI01 family has three representatives in worms and is named after one of the two homologues in budding yeast. There are also representatives from several Archaea species, but none from bacteria, making them a less attractive candidate as a progenitor to the protein kinase lineage.

Atypical Protein Kinases and Protein Kinase-Like Domains. Worms contain 26 kinase-like domains present in receptor guanylyl cyclases (there are 10 additional soluble guanylyl cyclases), and at least 7 diacylglycerol kinases, 7 choline/ethanolamine kinases, and 30 YLK1-related antibiotic resistance kinases. Each of these families contain short motifs that were recognized by our profile searches with low scoring E-values, but a priori would not be expected to function as protein kinases. Instead, the similarity could simply reflect the modular nature of protein evolution and the primal role of ATP binding in diverse phosphotransfer enzymes. However, two recent papers on a bacterial homologue of the YLK1 family suggests that the aminoglycoside phosphotransferases (APHs) are structurally and functionally related to protein kinases (28, 29). There are over 40 APHs identified from bacteria that are resistant to aminoglycosides such as kanamycin, gentamycin, or amikacin. The crystal structure of one well characterized APH reveals that it shares >40% structural identity with the two-lobed structure of the catalytic domain of cAMP-dependent protein kinase (PKA), including an N-terminal lobe composed of a five-stranded antiparallel β sheet and the core of the C-terminal lobe, including several invariant segments found in all protein kinases (29). APHs lack the GxGxxG normally present in the loop between β strands 1 and 2 but contain 7 of the 12 strictly conserved residues present in most protein kinases, including the HGDxxxN signature sequence in kinase subdomain VIB (29). Furthermore, Daigle et al. (30) have demonstrated that this APH also exhibits protein-serine/threonine kinase activity, suggesting that the worm YLK1-related molecules may indeed be functional protein kinases.

The eukaryotic lipid kinases (PI3Ks, PI4Ks, and PIPKs) also contain several short motifs similar to protein kinases but otherwise share minimal primary sequence similarity. However, once again, structural analysis of PIPKIIβ defines a conserved ATP-binding core that is strikingly similar to conventional protein kinases (31). Three residues are conserved among all of these enzymes, including (relative to the PKA sequence) Lys-72, which binds the α-phosphate of ATP, Asp-166, which is part of the HRDLK motif, and Asp-184, from the conserved Mg^{2+} or Mn^{2+} binding DFG motif (31). The worm genome contains 12 phosphatidylinositol kinases, including 3 PI3-kinases, 2 PI4-kinases, 3 PIP5-kinases, and 4 PI3-kinase-related kinases. The latter group has four mammalian members (DNA-PK, FRAP/TOR, ATM, and ATR), which have been shown to participate in the maintenance of genomic integrity in response to DNA damage and exhibit true protein kinase activity, raising the possibility that other PI-kinases may also act as protein kinases. Regardless of whether they have true protein kinase activity, PI3-kinases are tightly linked to protein kinase signaling, as evidenced by their involvement downstream of many growth factor receptors and as upstream activators of the cell survival response mediated by the AKT protein kinase.

There are several proteins with protein kinase activity that appear structurally unrelated to the eukaryotic protein kinases. These include Dictyostelium myosin heavy chain kinase A, Physarum polycephalum actin-fragmin kinase, the human A6 PTK, human BCR, mitochondrial pyruvate dehydrogenase and branched chain fatty acid dehydrogenase kinase, and the prokaryotic "histidine" protein kinase family. Worms lack representatives of the actin-fragmin kinases, BCR, and bacterial histidine kinases yet do contain a single representative of the other classes of atypical kinases and two homologues of the A6-related PTKs. The single worm orthologue of the Dictyostelium myosin heavy chain kinase A (32) proves to be the worm eukaryotic elongation factor 2 kinase (33). The slime mold, worm, and human eukaryotic elongation factor 2 kinase homologues have all been demonstrated to have protein kinase activity, yet they bear little resemblance to conventional protein kinases except for the presence of a putative GxGxxG ATP-binding motif (33).

The so-called histidine kinases are abundant in prokaryotes, with >20 representatives in E. coli, and have also been identified in yeast, molds, and plants. In response to external stimuli, these kinases act as part of two-component systems to regulate DNA replication, cell division, and differentiation through phosphorylation of an aspartate in the target protein (34). To date, no "histidine" kinases have been identified in metazoans, although mitochondrial pyruvate dehydrogenase (PDK) and branched chain α-ketoacid dehydrogenase kinase are related in sequence. PDK and branched chain α-ketoacid dehydrogenase kinase represent a unique family of atypical protein kinases involved in regulation of glycolysis, the citric acid cycle, and protein synthesis during protein malnutrition. Structurally, they conserve only the C-terminal portion of "histidine" kinases, including the G box regions. Branched chain α-ketoacid dehydrogenase kinase phosphorylates the E1α subunit of the branched chain α-ketoacid dehydrogenase complex on Ser-293, proving it to be a functional protein kinase (35). Although no bona fide "histidine" kinase has yet been identified in worms or humans, they do contain PDK homologues (one in worms and five in humans). However, the paucity of PDKs in worms makes it unlikely that they fill in for the absence of "histidine" kinases in metazoans. Instead, these signaling cascades have more likely been replaced by pathways initiated through RTKs.

Based on these examples of atypical protein kinases present in the worm genome, we predict additional worm protein kinases will be functionally identified that lack any of the obvious motifs conserved in the conventional members. Indeed, various biochemical data point to the existence of true histidine, lysine, and arginine kinases in metazoans, yet their structural identity remains a mystery.

Protein Phosphatases. Because of their important role in signal transduction, it is not surprising that the activity of protein kinases must be tightly regulated. This is accomplished through autoinhibition, autophosphorylation, transphosphorylation, dimerization, and cellular localization. Equally important is the role of protein phosphatases, which act to remove these regulatory phosphates from the protein kinase and its substrates. Because our analysis reveals worms to have a mature P.Tyr-signaling network, especially

when compared with the yeast genome, we surveyed the worm genome for protein-tyrosine phosphatases.

Our analysis reveals 83 conventional protein-tyrosine phosphatases (PTPs) plus 6 catalytic fragments and 12 additional fragments with high homology to the noncatalytic portion of other worm PTPs. In addition, there are 26 proteins classified as dual-specificity phosphatases related to MAPK phosphatases, CDC14, PRL, PIR1, CDC25, myotubularins, or PTEN lipid phosphatases. We also identify two SBF1- and one STYX-related proteins that are related to myotubularins and MAPK phosphatases yet lack the catalytic cysteine motif. These proteins are predicted to be catalytically inert yet may function as phosphoprotein-binding domains or anti-phosphatases (36). We also identify 11 inositol polyphosphate phosphatases and 65 serine-threonine phosphatases. Among the 83 PTPs, there are 57 cytoplasmic PTPs and 26 receptor-like PTPs, most of which are worm specific, lacking clear human orthologues. Exceptions include worm orthologues of the cytoplasmic PTPs; SHP2, MEG1, and MEG2, and the receptor PTPs; and PTPδ, PTPγ, PTPμ, PTPβ and IA2 (catalytically inactive). Overall, worms contain approximately the same number of tyrosine and dual-specificity protein kinases as they do tyrosine and dual-specificity protein phosphatases. This coordinate expansion in the eukaryotic lineage of both protein-tyrosine kinases and phosphatases emphasizes the biological need to maintain tight regulation of tyrosine phosphorylation. Because of the large numbers of worm-specific PTKs (FER and KIN-15 families) and worm-specific PTPs (89%, or 66 of 74), it is tempting to speculate that these unique enzymes may regulate each other's activity, or function in the same signaling pathways. Precedence for such specificity comes from genetic data indicating that the CLR-1 receptor PTP attenuates EGL-15, an FGFR orthologue, signaling in worms (37).

Conclusions. What does the worm genome sequence tell us about mammalian signal transduction? First, it has provided an ideal model to highlight the bioinformatics challenges that lie ahead with the human genome effort and allows us to test our analysis tools and database organization. Second, it lets us refine our expectations as to the diversity and absolute number of unique protein kinases that we can expect to find in the human genome. Based on our count of 493 (411 conventional and 82 PK-like proteins) worm kinases, minus the ≈197 kinases that appear to be worm-specific expansions of certain families such as the CK1, FER, and KIN-15 families, multiplied by the ≈4-fold greater number of genes in humans compared with worms, we predict the human genome to contain ≈1,100 protein kinases (PTKs and serine/threonine kinases). A similar extrapolation predicts ≈300 human protein phosphatases (PTPs, dual-specificity phosphatases, and serine-threonine phosphatases). Because our current count of human protein kinases and

phosphatases stands at ≈600 and 130, respectively, we still have about half the work ahead of us. However, recent claims predict the human genome may contain as many as 140,000 genes, compared with previous estimates of ≈80,000. Such calculations would result in a significant increase in our predictions of the total number of human protein kinases and phosphatases.

We may expect to see less evolutionary expansion of protein kinases families that serve elemental cellular functions such as cell cycle control and chromosome segregation, compared with processes involved in intercellular signaling or organogenesis. However, there is already evidence for at least a 2-fold expansion in the number of CDKs and "mitotic kinases" from worms to humans. Unlike expressed sequence tag data mining and PCR-based gene discovery approaches, genomic strategies do not bias against genes whose expression is tightly regulated in a cell-, developmental-, or disease-specific manner. This point is highlighted by the identification of 650 seven-transmembrane chemoreceptors in the worm genome (1), many of which may be expressed exclusively in single neurons. Because worms have only ≈302 neurons, compared with one trillion in humans, it would not be surprising to see this selectivity in cellular expression corroborated on mining the human genome. Indeed, because many of these novel protein kinases are likely to exhibit highly restricted expression, they may prove to be excellent targets for drug discovery in the battle against human disease.

The worm serves as a much simpler and tractable organism than humans for deciphering signaling cascades. Although their P.Tyr-signaling system is quite mature—based on the content of protein-tyrosine kinases, phosphatases, and adaptor molecules—they lack much of the molecular redundancy that exists in mammals, allowing the geneticist, biochemist, and cell biologist to more readily generate an "outline" of the signaling pathways that are conserved between worms and humans. The availability of the complete worm genome provides a unique opportunity to learn about human biology. Predicted orthologous pairs of human and worm genes can be targeted by using reverse genetic approaches to identify new signaling partners or biologic functions that can then be biochemically and functionally verified in mammals.

Although worms and humans have much in common, they also have obvious differences. Worms do not have limbs or bones, or a circulatory or immune system, and they eat only bacteria. Not surprisingly, they lack several protein kinases present in humans. Over the next 2 years, we should be better able to define which protein kinases are required for these specialized functions as the genome sequences of *Drosophila* and humans are completed. Identification and classification of the proteins present in each is just a first step toward understanding the biological complexity of life.

1. The *C. elegans* Sequencing Consortium (1998) *Science* **282**, 2012–2018.
2. Chervitz, S. A., Aravind, L., Sherlock, G., Ball, C. A., Koonin, E. V., Dwight, S. S., Harris, M. A., Dolinski, K., Mohr, S., Smith, T., *et al.* (1998) *Science* **282**, 2022–2028.
3. Clarke, N. D. & Berg, J. M. (1998) *Science* **282**, 2018–2022.
4. Ruvkun, G. & Hobert, O. (1998) *Science* **282**, 2033–2041.
5. Bingham, J., Plowman, G. D. & Sudarsanam, S. (1999) *J. Cell. Biochem.* in press.
6. Goffeau, A., Barrell, B. G., Bussey, H., Davis, R. W., Dujon, B., Feldmann, H., Galibert, F., Hoheisel, J. D., Jacq, C., Johnston, M., *et al.* (1996) *Science* **274**, 546, 563–567.
7. Hunter, T. & Plowman, G. D. (1997) *Trends Biochem. Sci.* **22**, 18–22.
8. Garret, J. M. (1997) *Mol. Microbiol.* **4**, 8098–8120.
9. Knippschild, U., Milne, D. M., Campbell, L. E., DeMaggio, A. J., Christenson, E., Hoekstra, M. F. & Meek, D. W. (1997) *Oncogene* **15**, 1727–1736.
10. Morgan, W. R. & Greenwald, I. (1993) *Mol. Cell. Biol.* **13**, 7133–7143.
11. Morgan, W. R. (1996) *Worm Breeder's Gazette* **14**, 27.
12. Murakami, S. & Johnson, T. E. (1998) *Curr. Biol.* **8**, 1091–1094.
13. Rosato R., Veltmaat, J. M., Groffen, J. & Heisterkamp, N. (1998) *Mol. Cell. Biol.* **18**, 5762–5770.
14. Metzstein, M. M., Stanfield, G. M. & Horvitz, H. R. (1998) *Trends Genet.* **14**, 410–416.
15. Hanks, S. K., Quinn, A. M. & Hunter, T. (1988) *Science* **241**, 42–52.
16. Downward, J. (1998) *Curr. Opin. Cell Biol.* **10**, 262–267.
17. Maehama, T. & Dixon, J. E. (1999) *Trends Cell Biol.* **9**, 125–128.
18. Bellanger, J. M., Lazaro, J. B., Diriong, S., Fernandez, A., Lamb, N. & Debant, A. (1998) *Oncogene* **16**, 147–152.
19. Sun, Z., Hsiao, J., Fay, D. S. & Stern, D. F. (1998) *Science* **281**, 272–274.
20. Matsuoka, S., Huang, M. & Elledge, S. J. (1998) *Science* **282**, 1893–1897.
21. Meneghini, M. D., Ishitani, T., Carter, J. C., Hisamoto, N., Ninomiya-Tsuji, J., Thorpe, C. J., Hamill, D. R., Matsumoto, K. & Bowerman, B. (1999) *Nature (London)* **399**, 793–797.
22. Korpelainen, E. I., Kärkkäinen, M., Gunji, Y., Vikkula, M. & Alitalo, K. (1999) *Oncogene* **18**, 1–8.
23. Hou, X. S., Melnick, M. B. & Perrimon, N. (1996) *Cell* **84**, 411–419.
24. Kawata, T., Shevchenko, A., Fukuzawa, M. Jermyn, D. A., Totty, N. F., Zhukovskaya, N. V., Sterling, A. E., Mann, M. & Williams, J. G. (1997) *Cell* **89**, 909–916.
25. Williams, J. C., Wierenga, R. K. & Saraste, M. (1998) *Trends Biochem. Sci.* **23**, 179–184.
26. Morin, P. J., Sparks, A. B., Korinek, V., Barker, N., Clevers, H., Vogelstein, B & Kinzler, K. W. (1997) *Science* **275**, 1787–1790.
27. Leonard, C. J., Aravind, L. & Koonin, E. V. (1998) *Genome Res.* **8**, 1038–1047.
28. Brasseur, G., Tron, G., Dujardin, G., Slonimski, P. P. & Brivet-Chevillotte. P. (1997) *Eur. J. Biochem.* **246**, 103–111.
29. Hon, W.-C., McKay, G. A., Thompson, P. R., Sweet, R. M., Yang, D. S. C., Wright, G. D. & Berghuis, A. M. (1997) *Cell* **89**, 887–895.
30. Daigle, D. M., McKay, G. A., Thompson, P. R. & Wright, G. D. (1999) *Chem. Biol.* **6**, 11–18.
31. Rao, V. D., Misra, S., Boronenkov, I. V., Anderson, R. A. & Hurley, J. H. (1998) *Cell* **94**, 829–839.
32. Cote, G. P., Luo, X., Murphy, M. B. & Egelhoff, T. T. (1997) *J. Biol. Chem.* **272**, 6846–6849.
33. Ryazanov, A. G., Ward, M. D., Mendola, C. E., Pavur, K. S., Dorovkov, M. V., Wiedmann, M., Erdjument-Bromage, H., Tempst, P., Parmer, T. G., Prostko, C. R., *et al.* (1997) *Proc. Natl. Acad. Sci. USA* **94**, 4884–4889.
34. Stock, J. (1999) *Curr. Biol.* **9**, R364–R367.
35. Harris, R. A., Hawes, J. W., Popov, K. M., Zhao, Y., Shimomura, Y., Sato, J., Jaskiewicz, J. & Hurley, T. D. (1997) *Adv. Enzyme Regul.* **37**, 271–293.
36. Hunter, T. (1998) *Nat. Genet.* **18**, 303–305.
37. Kokel, M., Borland, C. Z., DeLong, L., Horvitz, H. R. & Stern, M. J. (1998) *Genes Dev.* **12**, 1425–1427.

Exhibit 16

Analysis

Protein Kinases and Phosphatases in the *Drosophila* Genome

Deborah K. Morrison, Monica S. Murakami, and Vaughn Cleghon

Regulation of Cell Growth Laboratory, National Cancer Institute, Frederick, Maryland 21702

The reversible phosphorylation of proteins on serine, threonine, and tyrosine residues represents a fundamental strategy used by eukaryotic organisms to regulate a host of biological functions, including DNA replication, cell cycle progression, energy metabolism, and cell growth and differentiation. Levels of cellular protein phosphorylation are modulated both by protein kinases and phosphatases. Although the importance of kinases in this process has long been recognized, an appreciation for the complex and fundamental role of phosphatases is more recent. Through extensive biochemical and genetic analysis, we now know that pathways are not simply switched on with kinases and off with phosphatases. Rather, it is the balance of phosphorylation that is often critical. Protein phosphorylation can regulate enzyme function, mediate protein–protein interactions, alter subcellular localization, and control protein stability. Furthermore, kinases and phosphatases may work together to modulate the strength of a signal. Adding further complexity to this picture is the fact that both kinases and phosphatases can function in signaling networks where multiple kinases and phosphatases contribute to the outcome of a pathway. To fully understand this complex and essential regulatory process, the kinases and phosphatases mediating the changes in cellular phosphorylation must be identified and characterized.

A variety of approaches, including biochemical purification, gene isolation by homology, and genetic screens, have been successfully used for the identification of putative protein kinases and phosphatases. Now, the genomic sequencing of organisms promises to be a major contributor to this field. Valuable insight into these important enzymes has already emerged from the analysis of the yeast and worm genomes. In particular, genomic sequencing of *Saccharomyces cerevisiae* and *Caenorhabditis elegans* has revealed the kinase and phosphatase gene families that have arisen during the evolution of multicellular eukaryotes (Plowman et al., 1999). With the recent determination of the *Drosophila* sequence, we can now survey the genome of a second multicellular eukaryote for its repertoire of kinases and phosphatases. In this review, we will present our findings on the protein kinase and phosphatase gene families identified in the fly, together with an examination of the kinase/phosphatase signaling pathways functioning in flies, worms, and humans.

Address correspondence to Deborah Morrison, NCI-FCRDC, P.O. Box B, Frederick, MD 21702. Tel.: (301) 846-1733. Fax: (301) 846-1666. E-mail: morrisod@nciaxp.ncifcrf.gov

Vaughn Cleghon's current address is Beatson Institute CRC, Glasgow G61 1BD, UK. E-mail: v.cleghon@beatson.gla.ac.uk

Identification and Classification of Drosophila Protein Kinases and Phosphatases

Our survey of *Drosophila* protein kinases and phosphatases is based on the total set of predicted proteins that were identified in the *Drosophila* genome using automated gene predictor methods (Adams et al., 2000; available at http://www.celera.com). The 13,601 predicted fly proteins were surveyed for overall homology with known kinase and phosphatase sequences using BLASTP, and for the presence of polypeptide motifs using BLOCKS and InterPro databases (Rubin et al., 2000). Putative kinases and phosphatases identified by these means were further classified based on the presence of diagnostic amino acid residues in conserved motifs and by sequence similarities extending beyond conserved catalytic domains. Table I summarizes our survey of the *Drosophila* protein kinases and phosphatases. It is important to realize that this analysis represents the first tabulation of these enzymes in *Drosophila* and will be subject to revision as gaps in the genomic sequence are closed and methods for predicting and analyzing genes are improved. In particular, it is known that the Genie and Genscan programs used to annotate the fly genomic sequence make systematic errors with respect to intron–exon boundaries and gene borders, leading us to conclude that some kinase and phosphatase proteins may have been missed by these programs (Reese et al., 2000). These caveats notwithstanding, 251 kinases and 86 phosphatases were identified by our analysis of the predicted *Drosophila* protein set. Remarkably, more than half of these molecules had gone undetected in eight decades of *Drosophila* research.

Protein Kinases

Eukaryotic protein kinases are enzymes that catalyze the transfer of phosphate from ATP or GTP onto serine, threonine, or tyrosine residues of their appropriate substrates. They comprise a single protein superfamily having a common catalytic structure. However, these enzymes can be subdivided into distinct groups based on their structural and functional properties (Hanks and Hunter, 1995).

AGC Family

The AGC serine/threonine kinases function in many intracellular signaling pathways and were first classified based on their tendency to phosphorylate sites surrounded by basic amino acids. *Drosophila* contains ~30 AGC kinases, including members of the cyclic nucleotide-dependent ki-

Supplemental Material can be found at:
http://jcb.rupress.org/content/suppl/2000/08/01/150.2.F57.DC1.html

Table I. Summary of Protein Kinases and Phosphatases in Flies, Worms, and Humans   .

Group	Fly	Worm*	Humans*
Protein kinase			
AGC	30 (8)	30	100
CaMK	25 (13)	32	83
CKI	8 (6)	87	5
CMGC	24 (7)	42	62
STE	21 (12)	28	63
PTK	32 (8)	92	100
OPK	56 (28)	62	163
Atypical	3 (2)	4	11
Fragment/unknown	18		
Protein kinase like			
Gcyc	11 (6)	26	8
PIK	13 (8)	12	20
DAG	8 (5)	7	8
Choline K	2 (1)	7	2
Phosphatase			
STP	28 (14)	65	21
RPTP, CPTP, LMW-PTP	20 (12)	83	47
DSP	18 (11)	26	51
IPP	20 (18)	11	7

Fly numbers in parentheses represent the proteins newly identified by the fly genome project.
*These numbers are taken from the review by Plowman et al. (1999).

nases, protein kinase C (PKC),[1] AKT, NDR, MNK, MAST, ribosomal S6 kinase, and G protein–coupled receptor kinase families. The majority of the fly AGC kinases had been identified previously by molecular and genetic analysis; however, eight members were uncovered in the fly genome project. Interestingly, four of the new genes encode PKC or PKC-related proteins, including the first atypical PKC isoforms identified in *Drosophila*. Also identified by the fly genome project were additional PKA and PKG proteins, as well as kinases related to mammalian MAST205 and Citron.

CaMK Family

The CaMK serine/threonine kinases also tend to have substrate recognition motifs containing basic amino acids, and some but not all members of this family are regulated by calcium or calmodulin. Approximately 25 CaMKs are present in *Drosophila*, including representatives of the calcium/calmodulin-regulated kinase, SNF1/AMP-dependent kinase, EMK, CHK2, myosin light chain kinase (MLCK), phosphorylase kinase, death-associated protein kinase, and MAPKAP kinase families (the last four of which are found in *C. elegans* but not yeast). Like worms, flies do not encode a complete ortholog of the mammalian Trio kinase, but do have a protein that is related to the entire Trio regulatory domain. CaMK members revealed by the fly genome project include proteins related to calcium/calmodulin-regulated kinases, MLCK, EMK, and mammalian DRAK1. Of the 13 newly identified CaMKs, 6 be-

[1]*Abbreviations used in this paper:* CDK, cyclin-dependent kinase; CKI, casein kinase I; CTK, cytoplasmic tyrosine kinase; DSP, dual specificity phosphatase; LMW, low molecular weight; MKP, MAPK phosphatase; PKC, protein kinase C; PTP, protein tyrosine phosphatase; RTK, receptor tyrosine kinase; STP, serine/threonine protein phosphatase.

long to the EMK family, making this the largest CaMK group in flies. Mammalian and *C. elegans* EMK proteins have been implicated in the regulation of cell polarity and microtubule stability (Drewes et al., 1998).

Casein Kinase I Family

The casein kinase I (CKI) proteins originally were characterized as ubiquitous serine/threonine kinases with a preference for acidic substrates such as casein. Although members of this family were among the first kinases purified, elucidating their function and regulation has been difficult. Recently, however, CKI isoforms have been found to play a role in DNA repair and cell division (Gross and Anderson, 1998), in the Wnt signaling pathway (Peters et al., 1999), and in circadian rhythm regulation (Lowrey et al., 2000). *Drosophila* contains at least eight CKI proteins, only two of which were known previously. Intriguingly, CKI is one of the kinase families that is significantly expanded in the worm, with 87 members identified in *C. elegans* (Plowman et al., 1999). The biological significance of the worm-specific expansion is currently unknown.

CMGC Family

CMGC family members are primarily proline-directed serine/threonine kinases. The major subfamilies of this group play key roles in cell cycle regulation and intracellular signal transduction, and, not surprisingly, are conserved from yeast to humans. Approximately 24 CMGC kinases are found in *Drosophila*, including members of the cyclin-dependent kinase (CDK), CDC-like kinase (CLK), glycogen synthase kinase 3 (GSK3), and MAPK families. Although extensive genetic analysis had revealed many of the *Drosophila* CMGC kinases, seven novel proteins were uncovered by the fly genome project. These include additional CDK (CDK7-like, CDC2-related KKIALRE, CHED-related), GSK3, and MAPK (ERK7) members, as well as an RCK family member (MAK). Also uncovered in the fly genome were proteins related to the MP1 and JIP-1 scaffolding proteins. These molecules function to localize MAPK proteins with their upstream activators and provide signaling specificity (Whitmarsh and Davis, 1998). Although MAPK scaffolding proteins are present in yeast, they are structurally different from the ones found in flies, worms, and mammals, perhaps indicating the evolution of these molecules in multicellular eukaryotes.

STE Family

The STE family is composed of the STE7 (MEK), STE11 (MEKK), and STE20 (MEKKK) kinases that function upstream of MAPK proteins. *Drosophila* contains ~21 members of this family, only 9 of which were known previously. Remarkably, 9 members of the PAK/STE20 group were uncovered by the fly genome project, including proteins related to mammalian PAK3, GLK1, NIK, MST2, STLK3, TAO1, and CDC7. Although PAK proteins containing PH domains are found in yeast (Sells et al., 1999), no PH-domain-containing PAKs have been identified in higher eukaryotes and none are present in *Drosophila*. MEKK- and NEK-related kinases were also revealed by the genome project. It is worth noting that even with the discovery of additional MEK and MAPK proteins in the fly, *C. elegans*

contains over twice as many of these kinases, suggesting an expansion of MAPK signaling modules in the worm.

PTK Family

The PTK group consists of receptor (RTK) and cytoplasmic (CTK) tyrosine kinases. Although yeasts contain no conventional PTKs, 92 have been identified in the worm and ∼32 are present in the fly. A major function of PTKs is in intercellular communication, perhaps explaining why these enzymes have only been identified in multicellular eukaryotes. In comparison to *Drosophila*, the much larger number of PTKs found in *C. elegans* is due primarily to expansions of the worm-specific Kin-15/16 RTK and FER CTK families. The majority of the fly PTKs had been identified previously by genetic approaches, reflecting the involvement of these proteins in critical growth and developmental pathways. RTKs encoded in the fly genome include the fly-specific Torso and Sevenless kinases, as well as kinases related in sequence if not function to the mammalian EGFR, FGFR, insulin receptor, EPH, RET, ROR, RYK, ALK, and TRK kinases. Of the five newly identified RTKs, two are related to mammalian PDGFR/VEGFR, two are DDR receptors, and one shares homology with FGFR1. In the CTK group, fly members include the JAK, FAK, SYK/SHARK, ACK, ABL, and FPS kinases. Of the newly identified CTKs, one is related to mammalian ACK2 and one is an ortholog of CSK, a kinase that negatively regulates the activity of mammalian SRC kinases. Interestingly, several members of the PTK class are not found in worms, including representatives of the SYK, JAK, TRK, and RET families.

OPK Group

This group is comprised of other protein kinase (OPK) families that do not belong to the six major groups described above. It is the largest class of kinases found in flies and consists of both serine/threonine and dual specificity kinases. Approximately 56 of these enzymes are present in the fly genome, only half of which were known previously. Representatives of this group are extremely diverse and include members of the following families: Aurora, BUB1, CHK1, DYRK, WEE-1, PLK, EIF2, TGFβ, and activin receptor, TAK, IKK kinases, CKII, and RAF kinase. Notable in the novel group are additional BUB1 and TAK members and enzymes related to *C. elegans* UNC 51 and mammalian ALK3, DLK, GAK, MLK2, SRPK, IRE, ILK, TLK1, LIM-domain kinase, and LKB1/Peutz-Jeghers kinase.

Atypical, Lipid, and Unknown Kinases

Several protein groups that are structurally related to the eukaryotic protein kinases are also found in the *Drosophila* genome. These include the atypical kinases, guanylyl cyclases, and the eukaryotic lipid kinases. Flies contains at least three atypical kinase members, pyruvate dehydrogenase kinase, A6, and a newly identified BCR protein. Although worms lack BCR, they do contain a protein related to the atypical *Dictyostelium* myosin heavy chain kinase, which appears to be missing in flies. Also absent in both *Drosophila* and *C. elegans* are representatives of the classical prokaryotic histidine kinases. In the lipid kinase group,

Drosophila encodes at least 8 diacylglycerol kinases, 2 choline/ethanolamine kinases, and 13 phophatidylinositol kinases (PI3-, PI4-, PIP5,- and PIP3-related kinases), the majority of which were unknown previously. In mammalian cells, members of the PIP3-related kinase family participate in the cellular response to DNA damage and have authentic protein kinase activity (for review see Fruman et al., 1998). The fly genome project has revealed three kinases of this group, namely ATM, FRAP-related protein (FRP), and FRAP/TOR; however, as is true for worms, flies do not contain a DNA-PK. Finally, ∼18 proteins were identified that represent either partial kinase fragments or kinases with no significant homology to the groups listed above. Since errors have been identified in the transcript annotation of several protein kinases, such as the DDR receptors, Citron, and a PKC isoform, some of the partial kinase sequences may represent intact enzymes that have been improperly annotated. Further analysis will be required to confirm their identity.

Protein Phosphatases

Unlike protein kinases, which share a common catalytic structure, protein phosphatases have different basic structures, use distinct catalytic mechanisms, and comprise at least three separate protein families. Phosphatases are typically classified into two main groups, the serine/threonine protein phosphatases (STPs) and protein tyrosine phosphatases (PTPs).

STPs

STPs can be subdivided into the PPP and PPM families based on distinct amino acid sequences and crystal structures (for review see Cohen, 1997). Both families are widely distributed across phyla with representatives found in yeast, flies, worms, and mammals. Before the *Drosophila* sequencing project, almost all known fly STPs had been identified by molecular cloning approaches. Very few STPs have been isolated by genetic analysis, indicating that shared substrate specificity and/or functional redundancy may have prevented the recovery of such mutants. *Drosophila* contains ∼28 STPs, whereas >65 are encoded in the *C. elegans* genome. The increased number of worm STPs appears to be due to an expansion of the PPP family. Members of the PPP family, such as PP1, PP2A, and PP2B, have been implicated in numerous biological processes and signal transduction pathways. The diverse functions of this family are accomplished by a relatively small number of highly conserved catalytic subunits that complex with a wide variety of regulatory proteins, thus targeting the enzyme to specific intracellular locations and substrates. The *Drosophila* genome encodes ∼17 PPP catalytic proteins, 8 PP1-related enzymes (including PP1s, PPN, and PPY), 4 PP2A members (including PP2A, PP4, and PPV), 3 PP2B-like molecules, and 2 PP5 proteins. Additional PPP catalytic subunits uncovered by the fly genome project include members of the PP1, PP4, and PP2B groups. In regard to PPP regulatory subunits, *Drosophila* contains at least 3 PP1, 5 PP2A, and 2 PP2B proteins. However, because the regulatory subunits are so diverse, these numbers are likely to be low.

The PPM family includes PP2C and mitochondrial pyruvate dehydrogenase phosphatase. Due to their highly divergent primary sequences, few PPM members have been isolated by homology-based methods and none have been identified by genetic analysis. The only *Drosophila* PP2C protein that had been previously known was identified by genomic walking (Dick et al., 1997). Remarkably, the genome project has uncovered at least 11 new PP2C-related sequences, including one that closely resembles pyruvate dehydrogenase phosphatase. The biological function of the PPM family has been difficult to assess in mammalian cells due to the lack of specific inhibitors that target these enzymes. Recently, however, a PP2C protein has been found to dephosphorylate CDC2 on Thr161 in yeast (Cheng et al., 1999). Whether any of the PP2Cs perform a similar function in *Drosophila* waits to be determined.

PTPs

PTPs are found in all eukaryotic organisms, and are defined by the catalytic signature motif Cys-X5-Arg (for review see Neel and Tonks, 1997). The PTP superfamily consists of classical PTPs (RPTP, CPTP), dual specificity phosphatases (DSPs), and low molecular weight (LMW) PTPs. Approximately 38 PTPs are encoded in the fly genome, including representatives of each class. Again, many more PTPs are found in the worm (109 total). It is interesting to note that the expansion of serine/threonine and tyrosine kinase families in worms has been accompanied by a corresponding expansion of both serine/threonine and tyrosine phosphatases.

Members of the classical PTP family contain a conserved catalytic domain that is often fused to a large noncatalytic region. The PTP noncatalytic domains are quite diverse and can function to regulate enzyme activity and/or mediate protein interactions. Like PTKs, classical PTPs can be divided into two groups, receptor PTPs (RPTPs) and cytoplasmic PTPs (CPTPs). Genetic studies in *Drosophila* have been instrumental to our understanding of both groups. In particular, experiments in the fly were among the first to demonstrate the involvement of RPTPs in neuronal axon guidance (for review see Desai et al., 1997; den Hertog, 1999). *Drosophila* encodes ~8 RPTKs, at least 5 of which function in this capacity. Of the newly identified RPTPs, one is related to mammalian RPTP-κ and two share homology with RPTP-X/1A2, a type 1 transmembrane PTP implicated in nervous system development and insulin-mediated pancreatic function. In regard to the CPTP class, *Drosophila* studies on the CSW phosphatase were pivotal in demonstrating that a CPTP could function as a positive effector of cell signaling (Perkins et al., 1992). CSW is a member of the SH2-domain containing PTPs (SHP subclass). Mammals are known to have at least two SHPs, whereas no additional SHP proteins were found in *Drosophila*, indicating that flies, like worms, possess a single SHP molecule. Overall the fly genome encodes at least 5 CPTPs, namely CSW, PTP-ER, and newly identified CPTPs related to the mammalian MEG1, MEG2, and PTPD1 phosphatases. Finally, *Drosophila* contains four additional PTP-related proteins which are either difficult to classify or represent incomplete phosphatase fragments.

DSPs are a diverse collection of phosphatase subgroups

that share little sequence homology outside of the conserved Cys-X5-Arg motif with other DSP subgroups or with members of the larger PTP family. DSPs were originally characterized by their ability to dephosphorylate both serine/threonine and tyrosine residues; however, some of the DSP subgroups, namely PTEN and myotubularin, also possess lipid phosphatase activity (Maehama and Dixon, 1999). Approximately 18 DSPs are found in *Drosophila*, including representatives of the MAPK phosphatase (MKP), PTEN, nuclear prenylated PRL, myotubularin, PIR1, CDC14, and CDC25 phosphatase groups. Of the nine DSPs uncovered by the fly genome project, six belong to the MKP group, a remarkable finding considering the extraordinary effort spent studying MAPK pathways in *Drosophila*. Only Puckered, a negative regulator of the JNK pathway, previously had been identified by genetic techniques (Martin-Blanco et al., 1998). The failure of the new MKPs to be uncovered by genetic analysis may indicate that they participate in MAPK pathways controlling subtle or unappreciated phenotypes. Alternatively, their functions may have been obscured by redundancy within the MKP group or with other phosphatases. Additional DSPs revealed by the genome project include enzymes related to CDC14 and myotubularin. Interestingly, flies also contain three myotubularin-related sequences that lack the active site Cys and Arg residues. As has been suggested for similar mammalian myotubularin-related molecules, these proteins may function as antiphosphatases by binding to and protecting substrates from dephosphorylation by myotubularin or related phosphatase (Hunter, 1998; for review see Laporte et al., 1998).

LMW-PTPs are ~150-amino acid residue cytoplasmic enzymes that have been shown to possess tyrosine phosphatase activity (Ostanin et al., 1995). Other than a strictly conserved Cys-X5-Arg catalytic motif, LMW-PTPs bear little resemblance to the other PTP members. Mammalian LMW-PTPs have been implicated to function in EPH (Stein et al., 1998) and PDGF receptor signaling (Chiarugi et al., 2000); however, much remains to be learned regarding the biological activity of these enzymes. Although two putative LMW-PTPs are revealed by the *Drosophila* genome project, both predicted proteins are larger than would be expected (424 and 250 amino acids, respectively). The smaller protein contains a complete LMW-PTP domain but lacks the conserved Arg residue in the catalytic motif. Intriguingly, the larger protein has two complete LMW-PTP domains. Although the first domain has a mutation in the active site Cys residue and is likely to be inactive, the second domain contains an intact PTP catalytic motif and presumably has catalytic activity. If this protein is made in vivo, it would represent a new type of LMW-PTP having a tandem catalytic domain structure similar to that observed in many RPTPs. Whether this molecule is an authentic LMW-PTP and whether it has a human counterpart remains to be determined.

Lipid Phosphatases

Lipid inositol phosphatases play an important role in mediating the intracellular balance of second messenger phospholipids. *Drosophila* encodes approximately 20 inositol phosphatases (IPP), only 2 of which were known pre-

viously. Six inositol-1,4,5-triphosphase phosphatase–like enzymes are contained in the fly genome; yet as is true for worms, no ortholog of the mammalian SH2-domain–containing inositol 5' phosphatase (SHIP) appears to be present. *Drosophila* does encode eight PPAP enzymes, which dephosphorylate phosphatidic acid to generate di-acylglycerol. The prototype member of this class, Wunen, was first identified in a genetic screen for factors controlling germ cell migration in the early *Drosophila* embryo (Zhang et al., 1996). Related proteins were subsequently identified in yeast, worms, and mammals. Remarkably, the fly genome project reveals seven additional Wunen-like phosphatases. Also uncovered by the genome project are six members of the inositol monophosphate phosphatase (IMP) group. Both the Wunen-like and inositol monophosphate phosphatases are characterized by small tandem gene arrangements, suggesting a limited expansion of these phosphatase families in *Drosophila*. The large number of newly identified inositol phosphatases underscores the hitherto unappreciated importance of lipid phosphoregulation in the fly.

Comparative Analysis of Phosphorylation-dependent Signaling Pathways

With the completion of both the *Drosophila* and *C. ele-*

gans genome projects, together with our current knowledge of mammalian signaling pathways, we can begin to draw conclusions regarding the regulatory complexity of protein phosphorylation mechanisms across the evolutionary spectrum. For example, in flies, worms, and humans, there is a high degree of structural and functional conservation between the components of the RTK and stress-activated signaling pathways, with the major difference being the number of isoforms present for individual pathway members. In higher organisms, the number of isoforms is increased, presumably providing greater potential for tissue- or stage-specific functions, signaling cross-talk, and regulatory complexity (Fig. 1). Significantly, differences in phosphorylation-mediated signaling cascades between worms, flies and humans become apparent when examining the pathways involved in hematopoiesis and immunity. The JAK/STAT cascade, which has been implicated in hematopoiesis and cytokine signaling, is present in humans and flies. Worms, however, lack JAK kinases but do possess STAT proteins that are regulated by tyrosine phosphorylation. Like humans, flies also contain the Toll/IKK/NFκB pathway, which plays a role in the immune response to microbial organisms. No evidence of an inducible host defense system has been demonstrated in worms, consistent with the lack of this pathway in *C. elegans*. Also miss-

Human, Fly, Worm
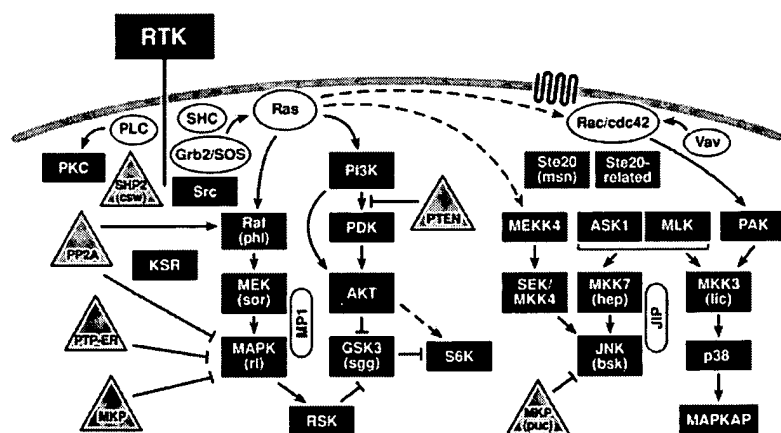
Human, Fly
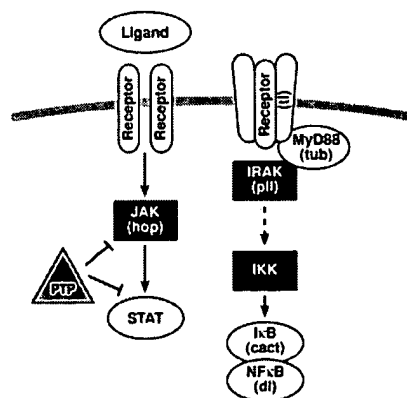
Human, Fly?
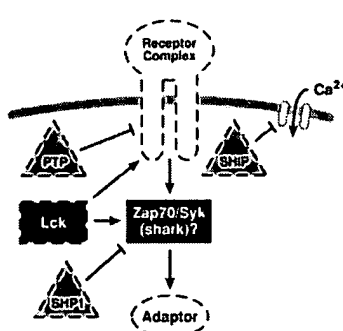
Figure 1. Comparison of the protein kinase/phosphatase signaling pathways in flies, worms, and humans (see text for description). Kinases are depicted as black rectangles, phosphatases are gray triangles, and other signaling components are in white. Shapes in dotted lines indicate mammalian proteins with no clear fly homologue; however, the function of these components in the pathway may be provided by other *Drosophila* proteins with related biochemical activities. *Drosophila* gene names are listed in parentheses.

ing in the worm are the SYK/ZAP70 kinases which play an important role in human T and B cell signaling. *Drosophila* may possess some form of this pathway as indicated by the presence of the fly SHARK kinase. The *Drosophila* SHARK kinase is a member of the SYK/ZAP70 family; however, it is most closely related to the HTK16 kinase of Hydra based on the presence of ANK repeats which are not found in any of the known mammalian SYK/ZAP70 family members (Chan et al., 1994; Ferrante et al. 1995). Exact homologues of proteins functioning with SYK/ZAP70 in the mammalian hematopoietic cascade, including the SLP-76, LAT, and BLNK adaptor proteins, the LCK and LYN kinases, and the SHP-1 and SHIP phosphatases were not revealed by the fly genome project; however, *Drosophila* proteins with related biological activities are found, namely SHP-2, inositol-1,4,5-triphosphate phosphatase, and other SRC-kinase members. Thus, further studies are required to determine whether a rudimentary form of the SYK/ZAP70 pathway does function in flies.

The completion of the *Drosophila* genome project also allows us to look globally at the pathways in which many of the newly identified fly enzymes may function. In particular, many of the proteins revealed in the *Drosophila* genome are orthologs of kinases and phosphatases known to function in the Rac/Rho/CDC42 signaling pathway (Citron, ACK2, MLK2, MEKK4, LIM-domain kinase, PAK/STE20, and DSPs members), in cell cycle regulation (CDK7, BUB1, NEK1, NEK2, CDC14, CDC7, and PP2C), and in pathways establishing asymmetry and cell polarity (LKB1, SLK1, and EMK kinases). Whether these enzymes went undetected for so many years because of functional redundancy or unappreciated phenotypes has yet to be determined.

In conclusion, ~251 protein kinases and 86 phosphatases have been identified in the *Drosophila* genome. Although the overall number of fly enzymes is lower than that found *C. elegans*, the difference is largely due to the worm-specific expansion of certain gene families. Interestingly, no large expansions or deletions of particular kinase or phosphatase gene families were uncovered by the *Drosophila* genome project. All of the previously known *Drosophila* kinases and phosphatases were detected in our analysis, confirming the relative completeness of the genome sequence data. Remarkably, almost 170 new protein kinases and phosphatases were identified by the fly genome project (Table I). The next challenge for scientists will be to determine the role of these enzymes in *Drosophila* development and physiology.

References

Adams, M.D., S.E. Celniker, R.A. Holt, C.A. Evans, J.D. Gocayne, P.G. Amanatides, S.E. Scherer, P.W. Li, R.A. Hoskins, R.F. Galle, et al. 2000. The genome sequence of *Drosophila melanogaster*. *Science*. 287:2185–2195.

Chan, T.A., C.A. Chu, K.A. Rauen, M. Kroiher, S.M. Tatarewicz, and R.E. Steele. 1994. Identification of a gene encoding a novel protein-tyrosine kinase containing SH2 domains and ankyrin-like repeats. *Oncogene*. 9:1253–1259.

Cheng, A., K.E. Ross, P. Kaldis, and M.J. Solomon. 1999. Dephosphorylation of cyclin-dependent kinases by type 2C protein phosphatases. *Genes Dev.* 13:2946–2957.

Chiarugi, P., P. Cirri, L. Taddei, E. Giannoni, G. Camici, G. Manao, G. Raugei, and G. Ramponi. 2000. The low M(r) protein-tyrosine phosphatase is involved in Rho-mediated cytoskeleton rearrangement after integrin and platelet-derived growth factor stimulation. *J. Biol. Chem.* 275:4640–4646.

Cohen, P.T.W. 1997. Novel protein serine/threonine phosphatases: variety is the spice of life. *Trends Biochem. Sci.* 22:245–251.

den Hertog, J. 1999. Protein-tyrosine phosphatases in development. *Mech. Dev.* 85:3–14.

Desai, C.J., Q. Sun, and K. Zinn. 1997. Tyrosine phosphorylation and axon guidance: of mice and flies. *Curr. Opin. Neurobiol.* 7:70–74.

Dick, T., S.M. Bahri, and W. Chia. 1997. *Drosophila* DPP2C1, a novel member of the protein phosphatase 2C (PP2C) family. *Gene.* 199:139–143.

Drewes, G., A. Ebneth, and E.M. Mandelkow. 1998. MAPs, MARKs, and microtubule dynamics. *Trends Biochem. Sci.* 23:307–311.

Ferrante, A.W., Jr., R. Reinke, and E.R. Stanley. 1995. Shark, a Src homology 2, ankyrin repeat, tyrosine kinase, is expressed on the apical surfaces of ectodermal epithelia. *Proc. Natl. Acad. Sci. USA.* 92:1911–1915.

Fruman, D.A., R.E. Meyers, and L.C. Cantley. 1998. Phosphoinositide kinases. *Annu. Rev. Biochem.* 67:481–507.

Gross, S.D., and R.A. Anderson. 1998. Casein kinase I: spatial organization and positioning of a multifunctional protein kinase family. *Cell Signal.* 10:699–711.

Hanks, S.K., and T. Hunter. 1995. Protein kinases 6. The eukaryotic protein kinase superfamily: kinase (catalytic) domain structure and classification. *FASEB (Fed. Am. Soc. Exp. Biol.) J.* 9:576–596.

Hunter, T. 1998. Anti-phosphatases take the stage. *Nat. Genet.* 18:303–305.

Laporte, J., F. Blondeau, A. Buj-Bello, D. Tentler, C. Kretz, N. Dahl, and J.L. Mandel. 1998. Characterization of the myotubularin dual specificity phosphatase gene family from yeast to human. *Hum. Mol. Genet.* 7:1703–1712.

Lowrey, P.L., K. Shimomura, M.P. Antoch, S. Yamazaki, P.D. Zemenides, M.R. Ralph, M. Menaker, and J.S. Takahashi. 2000. Positional syntenic cloning and functional characterization of the mammalian circadian mutation tau. *Science.* 288:483–492.

Maehama, T., and J.E. Dixon. 1999. PTEN: a tumour suppressor that functions as a phospholipid phosphatase. *Trends Cell Biol.* 9:125–128.

Martin-Blanco, E., A. Gampel, J. Ring, K. Virdee, N. Kirov, A.M. Tolkovsky, and A. Martinez-Arias. 1998. *puckered* encodes a phosphatase that mediates a feedback loop regulating JNK activity during dorsal closure in *Drosophila*. *Genes Dev.* 12:557–570.

Neel, B.G., and N.K. Tonks. 1997. Protein tyrosine phosphatases in signal transduction. *Curr. Opin. Cell Biol.* 9:193–204.

Ostanin, K., C. Pokalsky, S. Wang, and R.L. Van Etten. 1995. Cloning and characterization of a *Saccharomyces cerevisiae* gene encoding the low molecular weight protein-tyrosine phosphatase. *J. Biol. Chem.* 270:18491–18499.

Perkins, L.A., I. Larsen, and N. Perrimon. 1992. *corkscrew* encodes a putative tyrosine phosphatase that functions to transduce the terminal signal from the receptor tyrosine kinase *torso*. *Cell.* 70:225–236.

Peters, J.M., R.M. McKay, J.P. McKay, and J.M. Graff. 1999. Casein kinase I transduces Wnt signals. *Nature.* 401:345–350.

Plowman, G.D., S. Sudarsanam, J. Bingham, D. Whyte, and T. Hunter. 1999. The protein kinases of *Caenorhabditis elegans*: a model for signal transduction in multicellular organisms. *Proc. Natl. Acad. Sci. USA.* 96:13603–13610.

Reese, M.G., G. Hartzell, N.L. Harris, U. Ohler, J.F. Abril, and S.E. Lewis. 2000. Genome annotation assessment in *Drosophila melanogaster*. *Genome Res.* 10:483–501.

Rubin, G.M., M.D. Yandell, J.R. Wortman, G.L. Gabor Miklos, C.R. Nelson, I.K. Hariharan, M.E. Fortini, P.W. Li, R. Apweiler, W. Fleischmann, et al. 2000. Comparative genomics of the eukaryotes. *Science.* 287:2204–2215.

Sells, M.A., and J. Chernoff. 1997. Emerging from the PAK: the p21-activated protein kinase family. *Trends Cell Biol.* 7:162–167.

Stein, E., A.A. Lane, D.P. Cerretti, H.O. Schloecklmann, A.D. Schroff, R.L. Van Etten, and T.O. Daniel. 1998. Eph receptors discriminate specific ligand oligomers to determine alternative signaling complexes, attachment, and assembly responses. *Genes Dev.* 12:667–678.

Whitmarsh, A.J., and R.J. Davis. 1998. Structural organization of MAP-kinase signaling modules by scaffold proteins in yeast and mammals. *Trends Biochem. Sci.* 23:481–485.

Zhang, N., J. Zhang, Y. Cheng, and K. Howard. 1996. Identification and genetic analysis of *wunen*, a gene guiding *Drosophila melanogaster* germ cell migration. *Genetics.* 143:1231–1241.

Exhibit 17

ERK2 pathways, which contributes to the increased proliferative rate of tumor cells. For this reason, inhibitors of the ERK pathways are entering clinical trials as potential anti-cancer agents. In differentiated cells, ERKs have different roles and are involved in responses such as learning and memory in the central nervous system.

JNK1, JNK2, and JNK3

The JNKs were isolated and characterized as stress-activated protein kinases on the basis of their activation in response to inhibition of protein synthesis (8). The JNKs were then discovered to bind and phosphorylate the DNA binding protein c-Jun and increase its transcriptional activity. c-Jun is a component of the AP-1 transcription complex, which is an important · regulator of gene expression. AP-1 contributes to the control of many cytokine genes and is activated in response to environmental stress, radiation, and growth factors — all stimuli that activate JNKs. Regulation of the JNK pathway is extremely complex and is influenced by many MKKKs. As depicted in the STKE JNK Pathway Connections Map, there are 13 MKKKs that regulate the JNKs. This diversity of MKKKs allows a wide range of stimuli to activate this MAPK pathway. JNKs are important in controlling programmed cell death or apoptosis (9). The inhibition of JNKs enhances chemotherapy-induced inhibition of tumor cell growth, suggesting that JNKs may provide a molecular target for the treatment of cancer.

JNK inhibitors have also shown promise in animal models for the treatment of rheumatoid arthritis (10). The pharmaceutical industry is bringing JNK inhibitors into clinical trials for both diseases.

p38 Kinases

There are four p38 kinases: α, β, γ, and δ. The p38α enzyme is the most well characterized and is expressed in most cell types. The p38 kinases were first defined in a screen for drugs inhibiting tumor necrosis factor α-mediated inflammatory responses (11). The p38 MAPKs regulate the expression of many cytokines. p38 is activated in immune cells by inflammatory cytokines and has an important role in activation of the immune response. p38 MAPKs are activated by many other stimuli, including hormones, ligands for G protein–coupled receptors, and stresses such as osmotic shock and heat shock. Because the p38 MAPKs are key regulators of inflammatory cytokine expression, they appear to be involved in human diseases such as asthma and autoimmunity.

Recently, a major paradigm shift for MAPK regulation was developed for p38α. The p38α enzyme is activated by the protein TAB1 (12), but TAB1 is not a MKK. Rather, TAB1 appears to be an adaptor or scaffolding protein and has no known catalytic activity. This is the first demonstration that another mechanism exists for the regulation of MAPKs in addition to the MKKK-MKK-MAPK regulatory module. This important

observation indicates that other adaptor proteins should be scrutinized for potential roles in regulating MAPK activity.

The importance of MAPKs in controlling cellular responses to the environment and in regulating gene expression, cell growth, and apoptosis has made them a priority for research related to many human diseases. The ERK, JNK, and p38 pathways are all molecular targets for drug development, and inhibitors of MAPKs will undoubtedly be one of the next group of drugs developed for the treatment of human disease (13).

References and Notes

1. C. Widmann et al., Physiol. Rev. 79, 143 (1999).
2. L. B. Ray, T. W. Sturgill, Proc. Natl. Acad. Sci. U.S.A. 85, 3753 (1988).
3. G. L. Johnson, R. Lapadat, ERK Pathway, Science's STKE (Connections Map, as seen November 2002), http://stke.sciencemag.org/cgi/cm/stkecm; CMP_10705.
4. _____, JNK Pathway, Science's STKE (Connections Map, as seen November 2002), http://stke.sciencemag.org/cgi/cm/stkecm;CMP_10827.
5. _____, p38 Pathway, Science's STKE (Connections Map, as seen November 2002), http://stke.sciencemag.org/cgi/cm/stkecm;CMP_10958.
6. G. Zhou et al., Biol. Chem. 270, 12665 (1995).
7. M. K. Abe et al., Mol. Cell. Biol. 19, 1301 (1999).
8. J. M. Kyriakis et al., Nature 369, 156 (1994).
9. C. Tournier et al., Science 288, 870 (2000).
10. Z. Han et al., J. Clin. Invest. 108, 73 (2001).
11. J. C. Lee et al., Nature 372, 739 (1994).
12. B. Ge et al., Science 295, 1291 (2002).
13. J. M. English, M. H. Cobb, Trends Pharmacol. Sci. 23, 40 (2002).

The Protein Kinase Complement of the Human Genome

G. Manning,[1]* D. B. Whyte,[1] R. Martinez,[1] T. Hunter,[2] S. Sudarsanam[1,3]

We have catalogued the protein kinase complement of the human genome (the "kinome") using public and proprietary genomic, complementary DNA, and expressed sequence tag (EST) sequences. This provides a starting point for comprehensive analysis of protein phosphorylation in normal and disease states, as well as a detailed view of the current state of human genome analysis through a focus on one large gene family. We identify 518 putative protein kinase genes, of which 71 have not previously been reported or described as kinases, and we extend or correct the protein sequences of 56 more kinases. New genes include members of well-studied families as well as previously unidentified families, some of which are conserved in model organisms. Classification and comparison with model organism kinomes identified orthologous groups and highlighted expansions specific to human and other lineages. We also identified 106 protein kinase pseudogenes. Chromosomal mapping revealed several small clusters of kinase genes and revealed that 244 kinases map to disease loci or cancer amplicons.

Ever since the discovery nearly 50 years ago that reversible phosphorylation regulates the activity of glycogen phosphorylase, there has

been intense interest in the role of protein phosphorylation in regulating protein function. With the advent of DNA cloning and sequencing in

the mid-1970s, it rapidly became clear that a large family of eukaryotic protein kinases exists, and the burgeoning numbers of protein kinases led to the speculation that a vertebrate genome might encode as many as 1001 protein kinases (1). The near-completion of the human genome sequence now allows the identification of almost all human protein kinases. The total (518) is about half that predicted 15 years ago, but it is still a strikingly large number, constituting about 1.7% of all human genes.

Protein kinases mediate most of the signal transduction in eukaryotic cells; by modification of substrate activity, protein kinases also control many other cellular processes, including metabolism, transcription, cell cycle progression, cytoskeletal rearrangement and cell movement, apoptosis, and differentiation. Protein phosphorylation also plays a critical

[1]SUGEN Inc., 230 East Grand Avenue, South San Francisco, CA 94080, USA. [2]Salk Institute, 10010 North Torrey Pines Road, La Jolla, CA 92037, USA. [3]Genomics and Biotechnology, Pharmacia Corporation, 230 East Grand Avenue, South San Francisco, CA 94080, USA.

*To whom correspondence should be addressed. E-mail: gerard-manning@sugen.com

Table 1. Kinase distribution by major groups in human and model systems. A detailed classification is available in tables S1 and S6.

Group	Families	Subfamilies	Yeast kinases	Worm kinases	Fly kinases	Human kinases	Human pseudogenes	Novel human kinases
AGC	14	21	17	30	30	63	6	7
CAMK	17	33	21	46	32	74	39	10
CK1	3	5	4	85	10	12	5	2
CMGC	8	24	21	49	33	61	12	3
Other	37	39	38	67	45	83	21	23
STE	3	13	14	25	18	47	6	4
Tyrosine kinase	30	30	0	90	32	90	5	5
Tyrosine kinase–like	7	13	0	15	17	43	6	5
RGC	1	1	0	27	6	5	3	0
Atypical-PDHK	1	1	2	1	1	5	0	0
Atypical-Alpha	1	2	0	4	1	6	0	0
Atypical-RIO	1	3	2	3	3	3	1	2
Atypical-A6	1	1	1	2	1	2	2	0
Atypical-Other	7	7	2	1	2	9	0	4
Atypical-ABC1	1	1	3	3	3	5	0	5
Atypical-BRD	1	1	0	1	1	4	0	1
Atypical-PIKK	1	6	5	5	5	6	0	0
Total	134	201	130	454	240	518	106	71

role in intercellular communication during development, in physiological responses and in homeostasis, and in the functioning of the nervous and immune systems. Protein kinases are among the largest families of genes in eukaryotes (2–6) and have been intensively studied. As such, they made an attractive target for an initial in-depth analysis of the gene distribution in the draft human genome. Mutations and dysregulation of protein kinases play causal roles in human disease, affording the possibility of developing agonists and antagonists of these enzymes for use in disease therapy (7–9). A complete catalog of human protein kinases will aid in the discovery of human disease genes and in the development of therapeutics.

Comprehensive Discovery of Protein Kinase Genes

Most protein kinases belong to a single superfamily containing a eukaryotic protein kinase (ePK) catalytic domain. We set out to identify all sequenced human ePKs by searching every available human sequence source (public and Celera genomic databases, Incyte ESTs, in-house and GenBank cDNAs and ESTs) with a hidden Markov model (HMM) profile of the ePK domain (10). This profile is sensitive enough to detect short fragments of even very divergent kinases that have little similarity to any single known kinase. We extended these fragments to full-length gene predictions using a combination of EST and cDNA data, Genewise homology modeling, and Genscan ab initio gene prediction; more than 90% of the new and extended sequences were verified by cDNA cloning. We also identified 13 atypical protein kinase (aPK) families. These contain proteins reported to have biochemical kinase

activity, but which lack sequence similarity to the ePK domain, and their close homologs (10). Some aPKs have structural similarity to ePK domains (11). New aPKs were identified with the use of additional HMMs and Psi-Blast.

How Many Protein Kinases in the Genome?

We identified 478 human ePKs and 40 aPK genes (Table 1 and Fig. 1) (table S1). Of these 518 protein kinases, 24 are absent from the public Genpept database, and 47 more are

published only as hypothetical proteins or are not described as kinases. Many more are annotated only by automatic methods, or are fragmentary sequences and have not been individually studied. Most new kinases come from new and little-studied families, as targeted cloning has previously identified most members of well-known families. However, new members were found even in some of the best studied kinase families. One new member of the cyclin-dependent kinase (CDK) family was found: CDK11 is a close paralog of CDK8 (91% protein sequence identity for
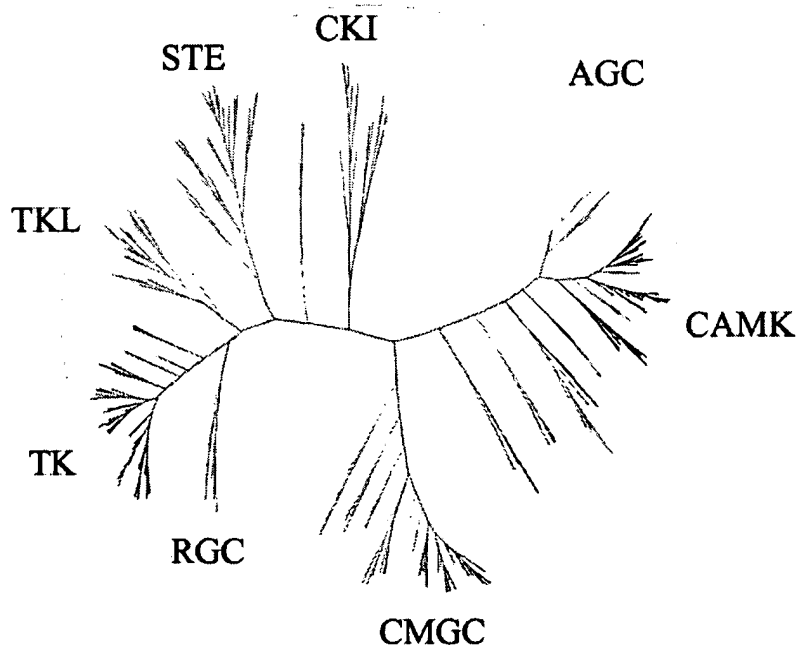
Fig. 1. Dendrogram of 491 ePK domains from 478 genes. Major groups (Table 1) are labeled and colored. For group-specific and comparative genomic trees, see www.kinase.com/human/kinome.

most of their length), a kinase that interacts with cyclin C and RNA polymerase II (*12*). A CDK11 ortholog exists in mouse, but fly (*Drosophila melanogaster*), worm (*Caenorhabditis elegans*), and yeast (*Saccharomyces cerevisiae*) have only a single member of this CDK8/CDK11 family. The Nek (NimA-related kinase) family is also thought to have a role in the cell cycle; we discovered four new Neks to bring the human total to 11 Nek kinases. Within the mitogen-activated protein kinase (MAPK) cascade, we found two new Ste11/MAP3K (MAP kinase kinase kinase) and two new Ste20/MAP4K (MAP kinase kinase kinase kinase) genes, all of which have restricted expression that may explain their failure to be previously cloned. For instance, only 14 ESTs are known from MAP3K8, and all but one derive from testis, lung, or brain libraries, indicating that these new genes may have evolved to mediate specialized roles in selected tissues.

Classification and Phylogeny of the Human Kinome

To compare related kinases in human and model organisms and to gain insights into kinase function and evolution, we classified all kinases into a hierarchy of groups, families, and subfamilies. This extends the Hanks and Hunter (*13*) human kinase classification of five broad groups, 44 families, and 51 subfamilies by adding four new groups, 90 families, and 145 subfamilies (Table 1 and Fig. 1) (table S1). Kinases were classified primarily by sequence comparison of their catalytic domains (*10*), aided by knowledge of sequence similarity and domain structure outside of the catalytic domains, known biological functions, and a similar classification of the yeast, worm, and fly kinomes (*4*).

Of the four new groups, STE consists of MAPK cascade families (Ste7/MAP2K, Ste11/MAP3K, and Ste20/MAP4K). The CK1 group contains CK1, TTBK (tau tubulin kinase), and VRK (vaccinia-related kinase) families. TKL (tyrosine kinase–like) is a diverse group of families that resemble both tyrosine and serine-threonine kinases. It consists of the MLK (mixed-lineage kinase), LISK (LIMK/TESK), IRAK [interleukin-1 (IL-1) receptor–associated kinase], Raf, RIPK [receptor-interacting protein kinase (RIP)], and STRK (activin and TGF-β receptors) families. Members of the RGC (receptor guanylate cyclase) group are also similar in domain sequence to tyrosine kinases.

Phylogenetic comparison of the human kinome with those of yeast, worm, and fly (*4*) confirms that most kinase families are shared among metazoans and defines classes that are expanded in each lineage. Of 189 subfamilies present in human, 51 are found in all four eukaryotic kinomes, and these presumably serve functions essential for the existence of a eukaryotic cell. An additional 93 subfamilies

are present in human, fly, and worm, implying that these evolved to fulfill distinct functions in early metazoan evolution. Comparison with the draft mouse genome indicates that more than 95% of human kinases have direct orthologs in mouse; additional orthologs may emerge as that genome sequence is completed.

The functions of human kinases can be inferred from family members in model organisms. For instance, the BRSK (brain-selective kinase) family has two uncharacterized human members that are selectively expressed in brain. They are orthologous to worm SAD-1, which has a role in presynaptic vesicle clustering (*14*), suggesting a conserved function. A highly conserved ascidian (chordate) homolog is also expressed in neural tissue and is asymmetrically localized to the posterior end of the embryo, suggesting a second role in embryonic axis determination (*15*). Conversely, we identified four families with orthologs in human, fly, and worm where no functional data are available for any member. Their phylogenetic distribution hints at roles fundamental to metazoan biology of which we are still ignorant.

The human genome has approximately twice as many kinases as those of fly or worm, after idiosyncratic worm-specific expansions are trimmed (*4*). Accordingly, most kinase families have twice as many human members as they have in worm or fly. However, the expansion is not uniform: 25 subfamilies—including CDK5, CDK9, and Erk7—have just one member in each organism, indicating critical unduplicated functions. Conversely, substantial human expansions occurred in several families, with the most striking example being Eph family receptor tyrosine kinases (RTKs), where there are 14 genes in human and only 1 in fly and worm (Table 2). These expanded families function predominantly in processes that are more advanced in human, such as the nervous and immune systems, angiogenesis, and hemopoiesis, as well as functions that are less obviously enhanced, such as apoptosis, MAPK signaling, calmodulin-dependent signaling, and epidermal growth factor (EGF) signaling.

Fourteen families are found only in human. The Tie family of RTKs are expressed in endothelial cells and function in angiogenesis, and the Axl RTKs (Axl, Mer, and Tyro3) function in both hemopoietic and neural tissues. The Trio and RIPK families have invertebrate homologs that lack kinase domains. They are involved in muscle function and apoptotic signaling via tumor necrosis factor (TNF), Fas, and NF-κB, respectively. Lmr, NKF3, NKF4, NKF5, and HUNK are novel families whose functions are largely unknown, and BCR, FAST, G11, H11, and DNAPK are atypical kinases.

The human expansions of many of these families can be traced both to large duplications of multigene loci ("paralogons") and to

local tandem duplications of smaller loci often containing just one gene. This supports recent findings that vertebrate genome complexity may derive from ancient large-scale duplications as well as a continuing series of smaller scale duplications (*16–18*). For instance, each of the four human epidermal growth factor receptors (EGFRs) maps close to one of the four HOX clusters, implying that the proposed double duplication of that cluster early in vertebrate evolution created the EGFR family from a single ancestral EGFR gene (*19*). Similarly, the eight genes of the VEGFR and PDGFR (vascular endothelial growth factor and platelet-derived growth factor receptors) families map to three of the four paraHOX clusters, and they probably derive from duplications of the single ancestral paraHOX locus as well as local duplications within the paraHOX loci (table S3). The common ancestry of PDGFR and VEGFR families is supported by the *Drosophila* kinome, which contains two genes whose sequences are intermediate between those two families (*4*).

We mapped all kinase genes to chromosomal loci to look for origins of kinase expansions and to link kinases with known disease loci. The map was created using the Celera and public genome assemblies and literature references (table S2). Although the overall kinase distribution is similar in density to that of other genes, many pairs of closely related genes from the same families map closer to each other than expected by chance, indicating that they may have arisen through local chromosomal duplications (table S3). Seven pairs are within 30 kb of each other, all in tandem orientation. Another six pairs are within 1 Mb of each other, and 15 more within 10 Mb. In all, 66 genes map unusually near to close paralogs, indicating that at least 6% of kinases may have arisen by local duplications. Most of these genes are from families that are highly expanded in human compared with worm and fly, further supporting a recent origin. The multigene duplications are thought to have arisen mostly during early vertebrate evolution, but some local duplications may also have happened at this time. For instance, the clustering of PDGFRβ and CSF-1 receptor (c-*fms*) genes is conserved in pufferfish (*20*).

Chromosomal Mapping and Disease

The knowledge of the exact chromosomal locations of genes afforded by the complete human genome assemblies is increasingly valuable in pinpointing candidate disease genes within loci that are associated with specific diseases. Comparison of the kinase chromosomal map with known disease loci indicates that 164 kinases map to amplicons seen frequently in tumors (*21*) and 80 kinases map to loci implicated in other major diseases (table

S2). Although each locus covers many genes, these data provide entry points for studying both the function of these kinases and their potential as the causative principle of these diseases. The role of kinases as biological control points and their tractability as drug targets make them attractive targets for disease therapy.

Catalytically Inactive Kinases

Several ePK domains are known to lack kinase activity experimentally, and these have been postulated to act as kinase substrates and scaffolds for assembly of signaling complexes (22–24). Our sequence analysis shows that 50 human kinase domains lack at least one of the conserved catalytic residues (Lys[30], Asp[125], and Asp[143]) (table S5) and are predicted to be enzymatically inactive. Twenty-eight inactive kinases belong to families where all members are inactive in human, fly, and worm, and even in yeast. Thus, surprisingly, nearly 10% of all kinase domains appear to lack catalytic activity. However, these domains are otherwise well conserved and are likely to maintain the typical kinase domain fold. This suggests that this domain can have generalized noncatalytic func-

tions; it is also possible that they use a modified catalytic mechanism that does not require these residues. This has been shown for the Wnk family, where Lys[13] is thought to replace Lys[30] in adenosine triphosphate (ATP) binding (25).

The 50 "inactive" kinase domains fall into three main categories. First are domains that may act as modulators of other catalytic domains. GCN2 and JAK (Janus kinase) family kinases have dual ePK domains, one of which is inactive and may regulate the active domain (26). Similarly, the inactive ePK domain of receptor guanylate cyclases (RGCs) is thought to regulate the activity of the neighboring guanylate cyclase domain, in a manner that is modulated by ATP binding and phosphorylation (27).

Second are other kinases with high similarity to the canonical ePK domain profile. These include the Ras pathway scaffold proteins KSR (kinase suppressor of Ras) (23) and the previously undescribed KSR2, titin, ILK (integrin-linked kinase), PSKH2 (protein serine kinase H2), and unpublished kinases from the STLK and Trbl families. The scaffold protein CASK (calcium/calmodulin-dependent serine kinase) contains an inactive protein kinase domain and

an inactive guanylate kinase domain, both of which act as protein-protein interaction domains (28, 29). This group also contains several RTKs where an inactive kinase may dimerize with and act as a substrate of another RTK: Ryk, CCK4, the ephrin receptors EphA10 and EphB6, and ErbB3 (24).

Third is a group whose members have very weak similarities to the kinase domain profile, and may have quite divergent functions. Of 37 "weak" kinase domains (whose kinase HMM E-value score is greater than 1e-30), 26 lack one or more catalytic residues. Note, however, that other weakly scoring kinases have been shown experimentally to have catalytic activity, including Bub1 (e-11 E value), VRK1 (e-10), PRPK (e-5), and haspin (e-3) (30–33).

Other Functional Domains in Protein Kinases

Most protein kinases act in a network of kinases and other signaling effectors, and are modulated by autophosphorylation and phosphorylation by other kinases. Other domains within these proteins regulate kinase activity, link to other signaling modules, or subcellu-

Table 2. Kinase families expanded in human relative to those in fly and worm. See table S6 for more details.

Function	Family	Human	Fly	Worm	Notes
Immunology, hemopoiesis, angiogenesis	JAK	4	1	0	Couple cytokine receptors to transcription
	PDGFR/VEGFR	8	2	0	Angiogenesis, vascular growth factor receptors
	Tec	5	1	0	Nonreceptor tyrosine kinase
	Src	11	2	3	Nonreceptor tyrosine kinase
	IRAK	4	1	1	IL-1 receptor–associated kinase
	Tie	2	0	0	Tie and Tek RTKs
	IKK	4	2	0	IκB kinase, NF-κB signaling
	RIPK	5	0	0	Receptor-interacting protein kinase, NF-κB signaling
	Axl	3	0	0	Immune system homeostasis
Neurobiology	Eph	14	1	1	Ephrin receptors
	Trk	3	0	0–1	Neurotrophin receptors
MAPK cascades	Ste11	9	2	2	(MAP3K)
	Ste20	31	13	12	(MAP4K)
	Ste7	8	4	10	(MAP2K) Has distinct worm-specific expansion
Apoptosis	DAPK	5	1	1	Death-associated protein kinase family
	RIPK	5	0	0	Transduces death signal from TNF-α receptor
	Lmr	3	0	0	Lmr1, aka apoptosis-associated tyrosine kinase (AATYK)
Calcium signaling	CaMK1	5	1	1	Calmodulin (CaM)–regulated kinases
	CaMK2	4	1	1	Calmodulin (CaM)–regulated kinases
EGF signaling	EGFR	4	1	1	Epidermal growth factor receptor family
	RSK/RSK	4	1	1	Ribosomal protein S6 kinases; RSK1-3 activated by MAPK in response to EGF
	Tao	3	1	1	Tao3 activated by EGFR
	Src	11	2	3	Src implicated in EGF signaling
	HUNK	1	0	0	Hormonally up-regulated Neu-associated kinase
Other	Trio	3	0	0	Fly and worm orthologs lack the kinase domain
	Trbl	3	1	0	Unpublished homologs of Drosophila trbl
	PDK	5	1	1	Mitochondrial pyruvate dehydrogenase kinases
	HIPK	4	1	1	Homeodomain-interacting protein kinases
	STKR	12	5	3	TGF-β, Activin receptors
	BRD	4	1	1	Bromodomain-containing atypical kinases
	Wnk	4	1	1	Implicated in hypertension
	NKF3	2	0	0	Uncharacterized (new kinase family 3)
	NKF4	2	0	0	Uncharacterized (new kinase family 4)
	NKF5	2	0	0	Uncharacterized (new kinase family 5)
	CDKL	5	1	1	Cyclin-dependent kinase-like

1915

larly localize the protein. We identified 83 additional types of domain present in 258 of the 518 kinases, using profiles from the Pfam HMM collection (Table 3). In general, members of the same kinase family have the same domain structure, but some domain shuffling is seen, where individual members of families have gained or lost a domain and so may have altered function. For instance, the death domain is found in all four IRAK kinases as well as in single members of the DAPK and RIPK families.

The most common domains mediate interactions with other signaling proteins: 24 kinases contain Src homology 2 (SH2) domains that bind to phosphotyrosine residues; other domains link to small guanosine triphosphatase (GTPase) signaling (38 kinases with RhoGEF, RhoGAP, RBD, PBD, RGS, CNH, HR1, or TBC do-

mains), lipid signaling (42 kinases with DAG_PE, C2, PX, or PH domains), and calcium signaling (28 kinases with CaM, IQ, or OPR/PB1 domains); target the protein to the cytoskeleton (seven kinases with spectrin, cofilin, myosin head, or FCH domains); or mediate interactions with other proteins (46 kinases: Death, SH3, SAM, LIM, or ankyrin domains) or RNA (three kinases with RRM, DSRM, and putative RNA binding Tudor domains). Most of the domains found in new or extended sequences are the same as those already seen in other family members, but some unpredicted domains are found, such as the previously unpublished leucine-rich repeat kinase (LRRK) family, containing arrays of leucine-rich repeats, as well as armadillo and ankyrin repeats.

Most of the 58 RTKs, 12 receptor serine-threonine kinases, and five receptor guanylate

cyclases also have recognizable ligand-binding and other extracellular domains, along with clear signal peptides and transmembrane regions. Several nonreceptor tyrosine kinases are also targeted to the membrane by lipidation or protein-protein interactions. Three kinases are targeted to the endoplasmic reticulum, five or six are likely to be mitochondrial, and most of the rest are thought to be cytoplasmic, nuclear, or both.

Two hundred and sixty kinases contain no additional Pfam domains. Many are small proteins containing little more than an ePK domain and may be controlled by additional regulatory subunits, such as cyclins, which control CDK activity. Others contain conserved sequences that have not yet been classified as domains and whose functions are unknown.

Thirteen kinases have dual ePK domains, in which both domains appear to be active [six ribosomal S6 kinase (RSK) family kinases and two Trio family kinases] or the second domain is inactive (the four JAK family kinases and GCN2). The two RSK domains are involved in a kinase relay: Erk phosphorylates and activates the CAMK-group domain of RSK2, leading to autophosphorylation on a linker region that then allows PDK1 to phosphorylate and activate the second AGC-group kinase domain (34).

Kinase Pseudogenes

The genome also contains many nonfunctional copies of kinase genes that are not expressed or encode degenerate, truncated proteins. These kinase pseudogenes are derived mostly from retroviral transposition and genomic duplications. Pseudogenes can confuse gene predictions, cross-hybridize with probes for functional genes, and contribute to disease by homologous recombination with their parental genes (35, 36). We identified 106 pseudogenes containing similarity to the ePK domain or to an aPK (table S4); several other pseudogene fragments that lack a kinase domain were found but are not included here. All but two pseudogenes have open reading frames (ORFs) interrupted by stop codons or frameshifts, which were verified by multiple independent sequence sources. These ORFs typically have high protein sequence similarity to a functional ("parent") kinase; most are partial gene fragments. The two putative pseudogenes with complete ORFs (CK2a-rs and STLK6-rs) lack introns and obvious promoters, are absent from EST databases, have >98.5% DNA sequence identity to their parents, and contain remnants of polyA tails in their genomic sequences. They are probably young processed pseudogenes whose sequences have not yet diverged.

Seventy-five kinase pseudogenes lack introns. Some are duplications of intronless genes

Table 3. Most common Pfam domains in protein kinases. See table S7 for a fuller listing.

Domain name	Number of genes	Number of domains	Function class
Protein kinase C terminal domain	44	44	Accessory domain
Immunoglobulin domain (Ig)	30	254	Extracellular, protein interactions
Fibronectin type III domain (FnIII)	28	194	Extracellular, protein interactions
SH2 domain	25	27	Adaptor: Binds phosphotyrosine
SH3 domain	27	28	Adaptor: Binds proline-rich motifs
PH domain	23	22	Signaling: phospholipid binding
Diacylglycerol binding (C1, DAG_PE)	23	33	Phospholipid binding
Calmodulin binding motif	23	25	Not in Pfam. From literature and sequence alignment
SAM domain (Sterile alpha motif)	15	16	Dimerization domain
Ephrin receptor ligand binding domain	14	14	Ligand binding
CNH domain	12	12	Cytoskeletal?
HEAT, armadillo/β-catenin repeats	10	27	Protein interaction
Activin receptor	11	11	Ligand binding
Ankyrin repeat (ANK)	9	59	Protein interaction
Regulator of G protein signaling (RGS)	7	7	GTPase interaction
PDZ/DHR/GLGF domain	7	7	Membrane targeting
Ubiquitin-associated domain A (UBA)	7	8	Protein degradation
Receptor L domain	7	14	Ligand binding
Furin-like cysteine rich region	7	21	Receptor dimerization?
p21-Rho-binding domain (PBD, CRIB)	9	9	GTPase interaction
Phosphatidylinositol 3'-kinase (PI3K)	6	6	Catalytic: Protein kinase
FAT	6	6	Accessory domain for PI3K
FATC	6	6	Accessory domain for PI3K
Alpha kinase	6	6	Catalytic: Atypical kinase
C2 domain	6	6	Ca²⁺, phospholipid binding
Guanylate cyclase catalytic domain	5	5	Catalytic: cGMP production
HSP90-like ATPase	5	5	Catalytic: Atypical kinase
ANF receptor	5	5	Ligand binding
Kinase-associated domain 1 (KA1)	5	5	Unknown
Bromodomain	8	13	Acetyl-lysine (chromatin) binding domain
HR1 repeat	5	13	GTPase interaction
Leucine-rich repeat	5	30	Ligand binding, protein interaction
ABC1 family	5	5	Catalytic: Atypical kinase
Death domain	6	6	Dimerization domain
BTK motif	4	4	Signaling
RhoGEF domain	4	5	GTPase interaction (guanine exchange factor)

or of single exons of larger genes, but most appear to derive from viral retrotransposition of a processed transcript. Additionally, some intron-containing pseudogenes such as AurAps2 contain some parental introns but lack others, and may result from retrotransposition of a partially spliced transcript.

Twenty-nine kinase pseudogenes contain clear introns and probably arose by genomic duplication. In some cases, these are part of a large duplicon (2, 5) containing multiple duplicated genes. Such cases include two p70 ribosomal protein S6 kinase (p70S6K) pseudogenes, which appear to arise from intrachromosomal duplications of the p70S6K locus. These duplications are 20 kb and 70 kb in length, and are 90 to 95% identical in DNA sequence to the original locus.

A few pseudogenes have no obvious human parent but have functional orthologs in rodents and probably indicate the decay of previously functional genes. They include the polo-like kinase SGK384ps, whose mouse ortholog is intact, and the human orthologs of rat guanylate cyclases CGD and KSGC.

Although pseudogenes appear to be evolutionary relicts, some may have some residual or cryptic function. Many pseudogenes are transcribed: 26 kinase pseudogenes are seen in cDNA and EST databases (table S4), some represented by as many as 50 ESTs.

The prevalence of pseudogenes varies greatly between kinase families (Table 1) (table S4). The MARK (microtubule affinity-regulating kinase) family kinases displays the largest ratio of pseudogenes to functional genes (28/4), followed by p70S6K (4/1), Erk3 (4/1), phosphorylase kinase γ1 (3/1), and casein kinase 1α (3/1). Frequent copying of a gene by retroviral insertion might indicate a functional role for the gene in retroviral function, but no viral function or source for MARK genes is yet known.

Comparison with Sequence Databases

We compared our nonredundant set of cloned and predicted kinase protein sequences with the published predictions from Celera and public genome projects (2, 5) and with a recent release of the public GenPept database (10). Figure 2 shows the extent to which the best match in each database agrees with our sequences. All three databases contain at least fragments of most kinases, but far fewer genes are in perfect agreement. In many cases the public sequences come from partial clones that lack the NH$_2$- or COOH-termini (43 and 15 genes, respectively), often from large-scale sequencing projects that do not individually annotate sequences. In other cases, the public sequence has overextended the true start site where upstream stop codons are absent. We used similarity to rodent orthologs to trim sequences to a strongly predicted translational start site in nine cases. Other discrepancies come from sequencing errors, alternative splicing, and sequencing of partially spliced

cDNAs. In all cases, our unique sequence is supported by strong sequence similarity to homologs or by cDNA cloning.

In some cases, our additional sequence greatly changes the predicted function of a gene, such as the addition of a predicted signal peptide to the Lmr1 tyrosine kinase; the previously published form of this gene (AATYK) was based on a cDNA lacking this domain, which created a cytoplasmic protein (37). We also identified full-length forms of two related new genes, Lmr2 and Lmr3, which together form a new family of predicted receptor tyrosine kinases with vestigial extracellular regions. Their biological roles are currently under investigation.

Gene predictions from the public genome project (Ensembl) and Celera differ from those we obtained largely as a result of misprediction of exon boundaries and splitting of single genes into multiple predicted genes. Ensembl incorporates public sequence data from RefSeq and Swiss-Prot, giving perfect agreement with our sequences for many genes. The distance between the GenPept and Ensembl traces in Fig. 2 indicates the extent of recent new sequence

tains multiple sequences for most kinases, many of which are partial fragments or contain multiple sequencing errors. It also contains chimeric genes such as the nonexistent zona pellucida kinase (38). The proliferation of different names for the same kinase adds to the problem of creating an accurate nonredundant list of kinases. Ensembl and Celera predictions include several pseudogenes (36 and 29, respectively), and also annotate as kinases a number of genes that are homologous to noncatalytic regulatory subunits of protein kinase complexes or to kinases other than protein kinases.

All 518 kinases are found in at least one of the expressed sequence databases (dbEST, Incyte, and GenBank cDNAs), indicating that all are genuine, transcribed genes. Many kinases are expressed in low amounts in a restricted distribution, so the presence of all kinases in EST or cDNA databases implies that these databases contain fragments of most human genes.

Summary

The sequencing of the human genome has provided a starting point for the identification of
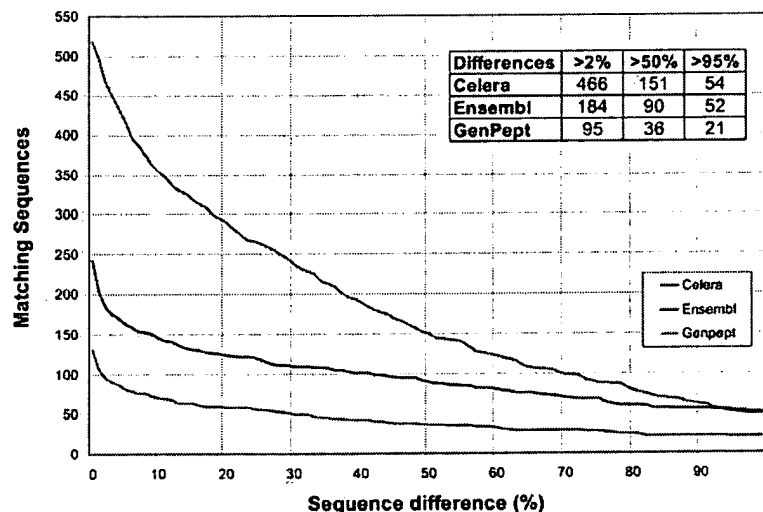
Fig. 2. Comparison of our kinase protein sequences with the best matches in Celera, Ensembl, and GenPept databases. Each point shows the number of genes for which the percentage difference between our sequence and the database is greater than the value indicated. Insert table indicates number of sequences where differences between our sequence and closest database match is >2%, >50%, or >95%.

publication from large-scale cDNA sequencing projects and individual cloning driven by genomic data. The Celera predictions were entirely computational, and so have very few perfect predictions. However, for genes not present in public databases, many Celera predictions agree better with our sequences than those from Ensembl (not shown).

A comparison with "known" protein kinases encounters several problems with over- and under-classification of genes as kinases, as well as with partial sequences. GenPept con-

most, if not all, human members of the eukaryotic protein kinase superfamily, and many atypical kinases. We used the published human genome sequences, combined with other sequence databases and directed cloning and sequencing of individual genes to discover, extend, or correct 125 kinase gene sequences, and define a nonredundant set of 518 human protein kinase genes. This set accounts for almost all human protein phosphorylation and collectively mediates most cellular signal transduction and many other processes. Comparative se-

quence analysis and mapping predict function and possible disease association for many kinases, and give clues to their evolutionary origin. Comprehensive kinome-scale approaches are now feasible, including RNA and protein expression profiling, and high-throughput functional assays using constitutively active and dominant-negative kinase constructs. These will facilitate the study of the role of kinases in a wide range of biological processes, and the development of selective inhibitors and activators for research and therapeutic purposes.

This large and well-curated sequence set also casts a light on the current state of human genome analysis. All 518 genes are covered by some EST sequence, and ~90% are present in gene predictions from the Celera and public genome databases, although those predictions are often fragmentary or inaccurate and are frequently misannotated (39).

References and Notes
1. T. Hunter, *Cell* **50**, 823 (1987).
2. E. S. Lander *et al.*, *Nature* **409**, 860 (2001).
3. G. M. Rubin *et al.*, *Science* **287**, 2204 (2000).
4. G. Manning, G. Plowman, T. Hunter, S. Sudarsanam, *Trends Biochem. Sci.* **27**, 514 (2002).
5. J. C. Venter *et al.*, *Science* **291**, 1304 (2001).
6. T. Hunter, G. D. Plowman, *Trends Biochem Sci.* **22**, 18 (1997).
7. P. Blume-Jensen, T. Hunter, *Nature* **411**, 355 (2001).
8. T. Hunter, *Cell* **100**, 113 (2000).
9. P. Cohen, *Nature Rev. Drug Discovery* **1**, 309 (2002).
10. See supporting data on *Science* Online and at www.kinase.com/human/kinome.
11. H. Yamaguchi, M. Matsushita, A. C. Naim, J. Kuriyan, *Mol. Cell* **7**, 1047 (2001).
12. P. Rickert *et al.*, *Oncogene* **12**, 2631 (1996).
13. S. K. Hanks, T. Hunter, *FASEB J.* **9**, 576 (1995).
14. J. G. Crump, M. Zhen, Y. Jin, C. I. Bargmann, *Neuron* **29**, 115 (2001).
15. Y. Sasakura, M. Ogasawara, K. W. Makabe, *Mech. Dev.* **76**, 161 (1998).
16. A. McLysaght, K. Hokamp, K. H. Wolfe, *Nature Genet.* **31**, 200 (2002).
17. X. Gu, Y. Wang, J. Gu, *Nature Genet.* **31**, 205 (2002).
18. L. Abi-Rached, A. Gilles, T. Shiina, P. Pontarotti, H. Inoko, *Nature Genet.* **31**, 100 (2002).
19. J. Spring, *Nature Genet.* **31**, 128 (2002).
20. G. F. How, B. Venkatesh, S. Brenner, *Genome Res.* **6**, 1185 (1996).
21. S. Knuutila *et al.*, *Am. J. Pathol.* **152**, 1107 (1998).
22. C. G. Zervas, N. H. Brown, *Curr. Biol.* **12**, R350 (2002).
23. D. K. Morrison, *J. Cell Sci.* **114**, 1609 (2001).
24. M. Kroiher, M. A. Miller, R. E. Steele, *Bioessays* **23**, 69 (2001).
25. B. Xu *et al.*, *J. Biol. Chem.* **275**, 16795 (2000).
26. M. Chen *et al.*, *Mol. Cell. Biol.* **20**, 947 (2000).
27. M. Chinkers, D. L. Garbers, *Science* **245**, 1392 (1989).
28. Y. Li, O. Spangenberg, I. Paarmann, M. Konrad, A. Lavie, *J. Biol. Chem.* **277**, 4159 (2002).
29. K. Tabuchi, T. Biederer, S. Butz, T. C. Sudhof, *J. Neurosci.* **22**, 4264 (2002).
30. H. Tanaka *et al.*, *J. Biol. Chem.* **274**, 17049 (1999).
31. S. Lopez-Borges, P. A. Lazo, *Oncogene* **19**, 3656 (2000).
32. T. W. Seeley, L. Wang, J. Y. Zhen, *Biochem. Biophys. Res. Commun.* **257**, 589 (1999).
33. Y. Abe *et al.*, *J. Biol. Chem.* **276**, 44003 (2001).
34. M. Frodin, C. J. Jensen, K. Merienne, S. Gammeltoft, *EMBO J.* **19**, 2924 (2000).
35. B. S. Emanuel, T. H. Shaikh, *Nature Rev. Genet.* **2**, 791 (2001).
36. B. Cormand, A. Diaz, D. Grinberg, A. Chabas, L. Vilageliu, *Blood Cells Mol. Dis.* **26**, 409 (2000).
37. E. Gaozza, S. J. Baker, R. K. Vora, E. P. Reddy, *Oncogene* **15**, 3127 (1997).
38. P. Bork, *Science* **271**, 1431 (1996).
39. We wish to thank the dozens of kinase researchers at SUGEN for their contributions to understanding the kinome at many levels. We particularly thank G. Plowman who guided the initial stages of the project, S. Caenepeel for extensive sequence analysis of kinases, and G. Charydczak for the computational support that made the genome mining possible. The SUGEN sequencing group provided cDNA confirmation of most predicted sequences. T.H. is a Frank and Else Schilling American Cancer Society Research Professor and serves on the Scientific Advisory Board of SUGEN.

Supporting Online Material
www.sciencemag.org/cgi/content/full/298/5600/1912/DC1
Materials and Methods
SOM Text
Tables S1 to S7
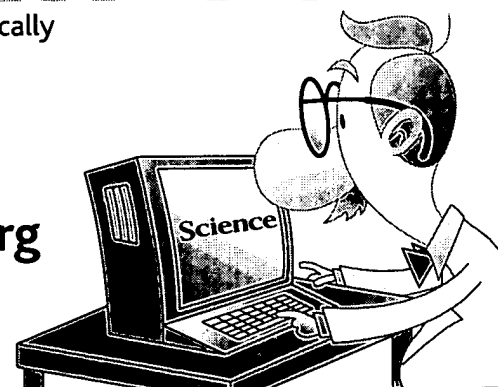
Cheaper Chips Find a Good Fit with Hit Validation
Drug Discovery & Development - February 03, 2005
 Protein microarrays promise to facilitate study of proteome interactions, and their use is growing despite proteins' inherent instability

Gina Shaw
Shaw is a contributing editor based in Marlboro, N.J.

In November 2004, protein array technology hit a milestone: Invitrogen Corp., Carlsbad, Calif., released the world's first commercially available high-density human protein microarray. The ProtoArray, which contains more than 1,800 unique human proteins, represents a cross-section of gene families including pharmaceutically relevant protein classes such as kinases and membrane-associated, cell-signaling, and metabolic proteins.

The pharmaceutical industry's demand for protein arrays has been high, and for good reason, says Steven Bodovitz, PhD, principal and cofounder of BioPerspectives, San Francisco, and an expert on protein biochips, protein biomarkers, and proteomics. "Mass [spectrometric] studies are very good at discovering potential targets and biomarkers, but they're not so good at following up with validation," he says. "You need to be able to take your initial findings and then see the protein change, and follow up and test it over and over again in different tissues, at different time points, under different conditions. To achieve that, you need a lower-cost, higher-throughput screening technology, which is a fantastic fit for the protein biochip."

Indeed, some experts have long been saying that protein arrays could potentially surpass DNA microarrays in their scientific impact. But moving from expectation to reality with these arrays has proven to be a longer journey than first imagined when protein arrays came on the scene in the late 1990s. "It was first thought that protein biochips would just be an extension of DNA microarrays, and that hasn't exactly panned out," says Bodovitz.

That's because proteins have proven to be much trickier to work with in array format than their genomic counterparts. First of all, there are issues of stability. Membrane proteins, for example, make up the majority of potential drug targets, but they're particularly challenging to stabilize. Then there's the choice of immobilization technique, which determines how well the target protein presents itself to the capture agent, and the problem of nonspecific binding. And of course, proteins are inherently unstable outside their natural habitat of living cells, making them much more challenging than DNA to tag and manipulate.

click the image to enlarge

Comparative analysis of normal and cancer sera using the Schleicher & Schuell Serum Biomarker Chip. Scatter plots of protein abundance ratios comparing serum from breast cancer patients with serum from healthy, age- and gender-matched individuals. Serum proteins were labeled with Biotin-ULS and Fluorescein-ULS, pooled, and probed against the Serum Biomarker Chip. (Source: Schleicher & Schuell)

Despite these challenges, though, the protein array market continues to grow. What was a $122 million market in 2002 will jump to $545 million by 2008, predicts an August 2004 report, *Protein Biochips: Parallelized Screening for High-Output Biology*. The report was released jointly by BioPerspectives; Bachmann Consulting in Nesoddtangen, Norway; and the NMI Natural and Medical Sciences Institute at the University of Tübingen, Germany. "The industry has begun to make the transition from a few years ago, where there were a lot of grandiose expectations, to very specific, aggressive approaches to developing protein biochips," says Bodovitz.

Although Invitrogen has lately been methodically gobbling up competitors and was the first company to offer a human protein biochip, the market for protein arrays is unlikely to be nailed down by a few leading vendors in the way that Affymetrix virtually cornered the market on DNA microarrays. "The genome is basically a limited set of information. Once you have a DNA microarray that covers the whole human genome, there is not a lot of room for something else," says Bodovitz. "That won't be the case at all with protein biochips. You have the capture and the interaction sides, which are very different technologies, and no one's yet covering the whole proteome, so no one big company is dominating."

Bright Shiny Beads
With all of the challenges inherent in developing solid-surface arrays that can hold thousands of proteins, all with different properties, it's little wonder that another approach—-protein-interaction assays in solution—-is also drawing attention. A leader in this market is Luminex Corp., Austin, Texas, whose bead-based xMAP system can provide up to 100 assay results from a single drop of sample. Like the planar arrays,

click the image to enlarge

Luminex's bead-based protein arrays attach reagents to the surface of color-coded beads, which can be studied in suspension at a rate of 100

A capture audience
What's the best approach to protein arrays? It depends on whom you ask, and what your goal is. Protein arrays can generally be broken down into two main categories: capture arrays and interaction arrays. Capture biochips use immobilized capture agents to capture target proteins, while interaction biochips have immobilized proteins or peptides which are used to identify functions or tease out interactions with other proteins or small molecules.

"For capture arrays, the big issue is content. In DNA microarrays, the oligo on the surface captures the target sequence and binds to it, but to capture a protein, you need a different, more complicated agent," says Bodovitz. "Historically, the problem has been that there aren't enough high-quality capture agents to populate a broad or high-density protein biochip platform. Antibodies don't work if they are denatured, for example."

Early solutions that appeared to have promise for rapidly generating research reagent antibodies don't appear to have panned out, so many companies are narrowing their sights to more focused content. "You organize a limited number of antibodies, say 30 or 40, maybe, sometimes more. The key is to give it a focus," says Bodovitz. "The most common focus is on cytokines. It's possible to cover a large number of the cytokine population on a biochip."

Luminex's protein arrays also attach capture reagents to the surface, in this case, a bead surface. Each bead is color-coded to distinguish what reaction is being demonstrated.

"The majority of protein array tools out there use a flat-surface, two-dimensional array. We try to mimic as well as possible what's actually happening in the cell. Biology doesn't happen in a tube," says James Jacobson, PhD, Luminex's vice president of research and development. "The beads allow a lot of things to happen that are advantageous. Because they're small and in suspension, we can take advantage of very favorable reaction kinetics. We also have a very high sample throughput; 100 results in a few seconds is pretty fast." Luminex has moved away from direct array sales and instead distributes them through a range of partnerships with companies like Bio-Rad Laboratories, Hercules, Calif., and Qiagen, Valencia, Calif., which provide add-on offerings like cytokine assays and assays for apoptotic markers and specific phospho-proteins, associated software, and reagents and kits that allow users to develop their own assays for the multiplexed environment.

That is an environment that interests many pharmaceutical scientists. "Multiplexing in protein arrays is promising, especially for cell signaling and signal transduction," says Eli Lilly's Myrtle Davis, DVM, PhD. "In those arenas, it's all about pathways, not just a single protein. We need to be able to identify a set of markers that indicate a particular modulation, not just one protein, and for that you need to see multiple protein interactions at one time."

targets within a few seconds. (Source: Luminex)

"The best arrays now only have 100 antibodies or less," says protein array expert Michael Snyder, director of the Yale Center for Genomics and Proteomics, New Haven, Conn. "I think it will take some new technologies to measure thousands of things using antibody arrays or other sorts of capture reagents. I don't think we know yet what are the best capture reagents, [be they] aptamers, monoclonal antibodies, or single-chain antibodies." Snyder thinks there are ways of improving the specificity of those reagents, ways that have yet to be explored.

One of the leaders in the capture array market is Schleicher & Schuell (S&S), Keene, N.H., which in June 2004 released its S&S Serum Biomarker Chip, the first high-density antibody chip specific to serum biomarkers related to human cancers of every major organ. The chip uses a cisplatin labeling chemistry that tags small molecules to serum proteins to label each sample with two different tags. The samples are then pooled and probed against the antibody microarray in a competitive binding fashion.

"The translational cancer research community didn't really have an easy-to-use tool that did not require specialized training, so the serum biomarker chip was probably the first product that addressed that need. Researchers can now identify a pattern of the abundance of protein in a diseased individual versus a matched healthy person, and, for example, identify a molecular signature that acts as a surrogate end point rather than a clinical end point, 12 to 16 weeks before you would see tumor response to a therapy," says business development manager Robert Negm, PhD.

As little as 8 mL can be processed to discriminate the abundance of more than 120 cancer serum biomarker proteins between two individuals. S&S has taken great pains to eliminate nonspecific binding, says Negm, using its FAST Quant TH1/TH2 assay (an alternative to microplate ELISAs for the assaying of multiple cytokines) to demonstrate that the antibody is not binding nonspecifically to another protein in the serum sample.

Still, their antibody-based nature remains the biggest limitation for capture arrays, says Myrtle Davis, DVM, PhD, senior research scientist at Eli Lilly and Co., Indianapolis. "Antibody-protein interaction is a wonderful thing that we can exploit as a means to pull out proteins from complex mixtures, but we do know that it's extremely limited."

At this point, she says, the quality of the protein array depends on the quality of the antibody. "One of the questions about antibody arrays that I always ask every vendor is this: if they use a chemistry to bind antibody to a solid phase, for example, how can they be sure that the antibody is bound in a conformation that allows it to capture antigens?" she asks. "You want to have some quality control on well-to-well variation."

Like Snyder, she thinks improved capture reagents are an important goal. "We need some other capture chemistries to be defined, so that these arrays can start to be more useful to the protein community."

Also on Davis' wish list is a protein binding technology that is less vendor-specific. "Reagents are often very specific to the vendor, for example, and we've found that when you start to employ some of these technologies, you're tied to the vendor," she says. "If a company goes under, you've set up an entire assay system around a technology you can no longer use."

Interaction in action
In terms of interaction arrays, Invitrogen appears to be unchallenged. Within the last year, Invitrogen acquired Protometrix, Branford, Conn., the developer of the Yeast ProtoArray, precursor to the Human ProtoArray. It also licensed rights to specific fields of use for more than 30 patents in the area of protein microarray development from Zyomyx Inc.

Interaction arrays are "the wild card," says Bodovitz. "If you talk to most protein biochemists, their reaction is, 'This can't possibly work.' Protein biochemistry is notoriously finicky, and people usually highly optimize any one reaction they're studying. Now you're talking about doing thousands of biochemical reactions on a chip surface? First, you're only studying them under one condition, plus immobilization could also have an effect. This means your data could only be applicable to one set of condition, and it may not be representative of anything."

He put this question to Protometrix scientists before the Invitrogen acquisition. "They countered that when they compared the reactions on the chip versus taking proteins

Table of No Content
If you had a protein chip with no proteins on it, then you'd have the ProteinChip system from Ciphergen Biosystems Inc., Fremont, Calif. The ProteinChip array consists of a variety of preactivated, chemically treated surfaces, designed for expression profiling when you're not sure just what protein you're looking for. Its main application is biomarker discovery and assay.

The ProteinChip, says Kate Gilbert, Ciphergen's director of marketing, is ideally suited to de novo discovery. "A researcher may be looking at a disease biomarker or efficacy biomarker and trying to predict response, and in many cases, may not be really sure what protein is going to prove to be a good marker," Gilbert says. "If you have antibodies down there, you're only going to find the proteins you have antibodies for. This approach, on the other hand, allows you to discover any type of protein, providing that it will bind under the chromatographic conditions you've selected."

The ProteinChip uses broad conditions in order to capture as much of the proteome as possible. The process is simple, designed as a benchtop system usable by individual researchers. First, a biological sample is put on the chip, and subpopulations of proteins are captured, retained and purified directly on the chip by affinity capture. The ProteinChip Reader uses a laser to desorb the retained proteins into a time-of-flight mass spectrometer; and accompanying software records and presents the molecular weight of the proteins found.

individually and doing it in solution, they got the same kind of results," he says. "Immobilization, at least, has been controlled for and apparently has very little impact on results. If you can make an initial discovery through a fast, very broad screen, and it holds up in subsequent assays, that's a very powerful method."

Pharmaceutical scientists are also attracted to Invitrogen's interaction arrays for specificity profiling. "The techniques currently used to assess antibody specificity are relatively crude. Western blots and such will largely characterize an antibody as specific or nonspecific, but they'll fail to identify exactly what the cross-reactivity is," says Predki. "With microarray experiments, in about a half a day, you can not only assess specificity but can immediately determine cross-reactive proteins."

"Our chips allow you to segregate proteins and give you the chance to see things you wouldn't see if you looked at one sample," says Gilbert.

http://www.dddmag.com/cheaper-chips-find-a-good-fit.aspx

Exhibit 19

new technology

Analysis of yeast protein kinases using protein chips

Heng Zhu[1], James F. Klemic[2,3], Swan Chang[2], Paul Bertone[1], Antonio Casamayor[1], Kathryn G. Klemic[4], David Smith[1], Mark Gerstein[5], Mark A. Reed[2,3] & Michael Snyder[1,5]

We have developed a novel protein chip technology that allows the high-throughput analysis of bio-chemical activities, and used this approach to analyse nearly all of the protein kinases from *Saccharomyces cerevisiae*. Protein chips are disposable arrays of microwells in silicone elastomer sheets placed on top of microscope slides. The high density and small size of the wells allows for high-throughput batch processing and simultaneous analysis of many individual samples. Only small amounts of protein are required. Of 122 known and predicted yeast protein kinases, 119 were overexpressed and analysed using 17 different substrates and protein chips. We found many novel activities and that a large number of protein kinases are capable of phosphorylating tyrosine. The tyrosine phosphorylating enzymes often share common amino acid residues that lie near the catalytic region. Thus, our study identified a number of novel features of protein kinases and demonstrates that protein chip technology is useful for high-throughput screening of protein biochemical activity.

Introduction

The sequencing of entire genomes has resulted in the identification of large numbers of novel ORFs. The challenge ahead is to gain information about the function of identified genes[1,2]. Currently, significant effort is devoted to understanding gene function by mRNA expression patterns and by gene disruption phenotypes[3,4]. Important advances in this effort have been possible, in part, by the ability to analyse thousands of gene sequences in a single experiment using gene chip technology. Much information about gene function comes from the analysis of the biochemical activities of the encoded protein. Currently, these types of analyses are done by individual investigators studying a single protein at a time. This can be time consuming because it can take years to purify and identify a protein on the basis of its biochemical activity. The availability of an entire genome sequence makes it possible to perform biochemical assays on every protein encoded by the genome. As such, it would be extremely powerful to analyse hundreds or thousands of protein samples using a single protein chip. Such approaches lend themselves well to high-throughput experiments in which large amounts of data can be generated and analysed.

Several groups have devised methods for expressing large numbers of proteins with potential utility for biochemical genomics in *S. cerevisiae*. InVitrogen has cloned ORFs into an expression vector that uses the *GAL* promotor and fuses the protein to a HISX6 tag; thus far they have prepared and confirmed expression of approximately 2,000 yeast protein fusions[5]. Using a recombination strategy, Eric Phizicky's group has cloned approximately 85% of the yeast ORFs into a vector that produces GST fusion proteins under the control of the *CUP1* promotor (inducible by copper[6]). Using a pooling strategy, they identified the gene encoding several important biochemical activities (for example, phosphodiesterase and Appr-1''-P-processing activities). Strategies to analyse large numbers of individual protein samples have not been described.

We have also overproduced yeast proteins as GST fusions and

developed a protein chip technology suitable for rapidly analysing large numbers of samples; this approach was applied to the analysis of nearly all yeast protein kinases. The yeast genome has been sequenced and contains approximately 6,200 ORFs greater than 100 codons in length. Of these, 122 are predicted to encode protein kinases, and 24 of these protein kinase genes have not been studied previously[7]. Except for two histidine protein kinases, all of the yeast protein kinases are members of the Ser/Thr family; tyrosine kinase family members do not exist, although seven protein kinases that phosphorylate serine/threonine and tyrosine have been reported[7].

Here we overexpress nearly all (119) of the yeast protein kinases and used a novel protein chip technology to analyse their specificity using 17 different substrates. We find that 32 kinases preferentially phosphorylate one or two substrates, and 27 kinases readily phosphorylate poly(Tyr-Glu), suggesting that there are many more potential tyrosine kinases than were known previously. Correlation of functional specificity with amino acid sequence information reveals that the kinases that use poly(Tyr-Glu) as a substrate contain amino acids near the catalytic region that are distinct from those that do not. We expect this technology to be valuable for the analysis of entire proteomes and the information to be very valuable to researchers studying kinase-substrate reactions.

Results
Yeast kinase cloning and protein purification

Using a recombination-directed cloning strategy[8], we cloned the entire coding regions of 122 yeast protein kinase genes in a high-copy expression vector (pEG(KG)) that produces GST fusion proteins under the control of the galactose-inducible *GAL1* promoter[9] (Fig. 1a). GST::kinase constructs were rescued into *Escherichia coli*, and sequences at the 5' end of each construct were determined. We successfully cloned 119 of the protein kinase genes in-frame. The three kinase genes that we did not clone were very large (4.5–8.4 kb).

Departments of [1]Molecular, Cellular, and Developmental Biology, [2]Electrical Engineering, [3]Applied Physics, [4]Cellular and Molecular Physiology, and [5]Molecular Biophysics and Biochemistry, Yale University, New Haven, Connecticut, USA. Correspondence should be addressed to M.S. (e-mail: michael.snyder@yale.edu).

a

b

radioisotope labelling (^{33}P), the kinase assays described below and manual loading, we tested a variety of microarray configurations and found that the following chips produced the best results: round wells 1.4 mm in diameter and 300 μm deep (approximately 300 nl), in a 10×14 rectangular array configuration with a 1.8 mm pitch. We then made a master mold of 12

The GST::kinase fusion proteins were overproduced in yeast and purified from 50-ml cultures using glutathione beads and standard protocols[10]. For the case of Hog1p, in the last five minutes of induction the yeast cells were treated with high salt to activate the enzyme; for the rest of the kinases, synthetic media (URA⁻/raffinose) was used. Immunoblot analysis of all 119 fusions using anti-GST antibodies revealed that 105 of the yeast strains produced detectable GST::fusion proteins; in most cases the fusions were full length. Up to 1 μg of fusion protein per millilitre of starting culture was obtained (Fig. 1*b*), but we failed to detect 14 of 119 GST::kinase samples by immunoblotting analysis, despite repeated attempts. Presumably, these proteins are not stably overproduced in the *pep4* protease-deficient strain used, or these proteins may form insoluble aggregates that do not purify using our procedures. Although this procedure was successful, purification of GST fusion proteins using 50-ml cultures is time consuming and is not applicable for preparing thousands of samples. Therefore, we have developed a procedure for purifying proteins in a 96-well format. Using this procedure, we prepared and purified 119 GST fusions in 6 hours with approximately twofold higher yields per millilitre of starting culture relative to the 50-ml method.

Protein chip design

We developed protein chips to conduct high-throughput biochemical assays of these 119 protein kinases (Fig. 2). These chips consist of an array of microwells in a disposable silicone elastomer, poly(dimethylsiloxane) (PDMS; ref. 10). Microwell arrays allow small volumes of different analytes to be densely packed on a single chip, yet remain physically segregated during subsequent batch processing. Proteins were covalently attached to the wells using a crosslinker 3-glycidoxypropyltrimethoxysilane[11] (GPTS). Up to 8×10⁻⁹ μg/μm² of protein can be attached to the surface.

For the purposes of the protein kinase assays described here, we configured the protein chip technology to be compatible with standard sample handling and recording equipment. Using

of these arrays and repeatedly cast microarrays for the protein kinase analysis. Chips were placed atop microscope slides for handling purposes (Fig. 2*a*); the arrays covered slightly more than one-third of a standard microscope slide and we typically used two arrays per slide (Fig. 2*b*). Although we used a manual pipette method to place proteins in each well, automated techniques may also be used. In addition, this protein chip configuration may also be used with other tagging methods such as fluorescent antibodies.

Large-scale kinase assays using protein chips

All 119 GST:protein kinases were tested for *in vitro* kinase activity[12] in 17 different assays using ^{33}Pγ-ATP and the following 17 substrates: (i) the kinases themselves (autophosphorylation); (ii) bovine histone H1 (a common kinase substrate); (iii) bovine casein (a common substrate); (iv) myelin basic protein (a common substrate); (v) Axl2 carboxy terminus-GST (Axl2 is a transmembrane phosphoprotein involved in budding[13]); (vi) Rad9 (a phosphoprotein involved in the DNA damage checkpoint[14]); (vii) Gic2 (a phosphoprotein involved in budding[15]); (viii) Red1 (a meiotic phosphoprotein important for chromosome synapsis[16]); (ix) Mek1 (a meiotic protein kinase important for chromosome synapsis[17]); (x) Poly(tyrosine-glutamate 1:4) (poly (Tyr-Glu); a tyrosine kinase substrate[18]); (xi) Ptk2 (a small-molecule transport protein[19]); (xii) Hsl1 (a protein kinase involved in cell cycle regulation[20]); (xiii) Swi6 (a phosphotranscription factor involved in G1/S control[21]); (xiv) Tub4 (a protein involved in microtubule nucleation[22]); (xv) Hog1 (a protein kinase involved in osmoregulation[23]); (xvi) Hog1 (an inactive form of the kinase); and (xvii) GST (a control). For the autophosphorylation assay, the kinases were directly adhered to the treated PDMS wells and ^{33}Pγ-ATP was added; for substrate reactions, the substrates were bound to the wells, and then kinases and ^{33}Pγ-ATP were added. After the reactions were completed, the slides were washed and the phosphorylation signals were acquired and quantified using a high-resolution phosphoimager (Fig. 3). To identify kinase activities, the quantified signals were converted into

Fig. 2 Protein chip fabrication and kinase assays. *a*, Kinase activities were detected using protein chips. PDMS was poured over the acrylic mold. After curing, the chip containing the wells was peeled away and mounted on a glass slide. The next step included modification of the surface and then attachment of proteins to the wells. Wells were blocked with 1% BSA before kinase, $^{33}P\gamma$-ATP and buffer were added. After incubation for 30 min at 30 °C, the chips were washed extensively and exposed to both X-ray film and a phosphoimager, which has a resolution of 50 µm and is quantitative. For 12 substrates each kinase assay was repeated at least twice; for the remaining 5 the assays were performed once. *b*, An enlarged picture of the protein chip.

a

etched mold → pour PDMS on surface → microwells

kinase assay ← substrate attached ← activated by GPTS
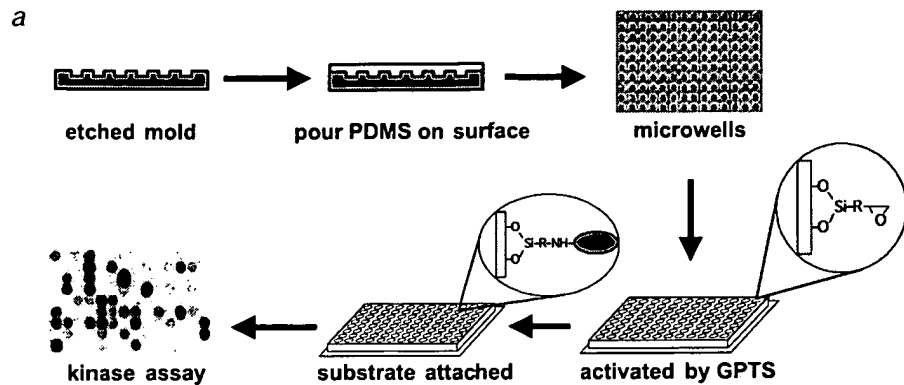
fold increases relative to GST controls and plotted for further analysis (Fig. 4a).

Most (112/119; 94%) kinases exhibited activity fivefold or greater over background for at least one substrate (Fig. 4a). As expected, Hrr25p, Pbs2p and Mek1p phosphorylated their known substrates[24–26], Swi6p (400-fold higher than the GST control), Hog1p (10-fold higher) and Red1p (10-fold higher), respectively. Using this assay, we

b

28 mm
18 mm
1.4 mm

found that 18 of 24 predicted protein kinases that have not been previously studied phosphorylate one or more substrates. Several unconventional kinases[7], including the histidine kinase YIL042c and phospholipid kinase Mec1p, phosphorylate protein substrates in *trans*.

To determine substrate specificity, the activity of a particular kinase was further normalized against the average of its activity against all substrates (Fig. 4b; all data are available at http://bioinfo.mbb.yale.edu/genome/yeast/chip). We found that 32

kinases had substrate specificity on a particular substrate with specificity index (SI) equal or higher than 2, and, reciprocally, most substrates are preferentially phosphorylated by a particular protein kinase or set of kinases. For example, the preferred substrates for YIL042C and Mec1p were Swi6p and Axl2p. The C terminus of Axl2, a protein involved in yeast cell budding, is also preferentially phosphorylated by Dbf20p, Kin2p, Yak1p and Ste20p relative to other proteins. Previous studies found that Ste20p was localized at the tip of emerging buds similar to Axl2p, and a *ste20Δcla4^{ts}* mutant is unable to bud or form fully polarized actin patches or cables[27]. Another example is the phosphoprotein Gic2, which is also involved in budding[15]. Ste20p and Skm1p strongly phosphorylate Gic2p (Fig. 4b). Previous studies suggested that Cdc42p interacts with Gic2p, Cla4p (ref. 28), Ste20p and Skm1p. Our results raise the possibility that Cdc42p may function to promote the phosphorylation of Gic2p by recruiting Ste20p and/or Skm1p.

autophos casein histone H1 GIC2

MBP MEK1 poly(E-Y) RED1

RAD9 SWI6 TUB4 PTK2
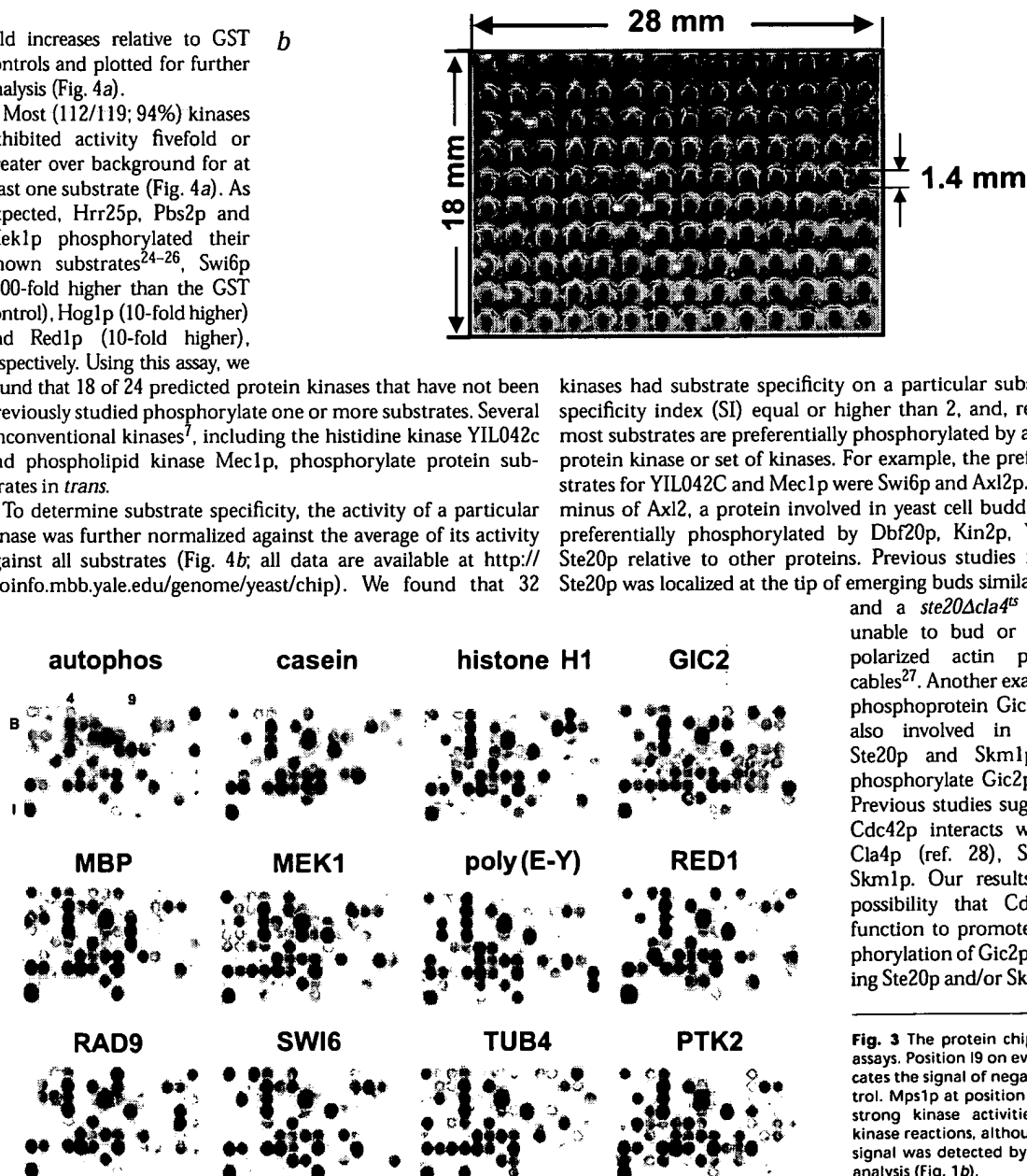
Fig. 3 The protein chip and kinase assays. Position I9 on every chip indicates the signal of negative GST control. Mps1p at position B4 exhibited strong kinase activities in all 12 kinase reactions, although no visible signal was detected by immunoblot analysis (Fig. 1b).

new technology

Many yeast kinases phosphorylate poly(Tyr-Glu)

On the basis of sequence analysis, all but two yeast protein kinases belong to the Ser/Thr family of protein kinases; the two exceptions are members of the histidine kinase family. Proteins of the conventional tyrosine kinase sequence family are lacking. At the time we started our study, however, seven protein kinases (Mps1, Rad53, Swe1, Ime2, Ste7, Hrr25 and Mck1) were reported to phosphorylate tyrosine[18]. We confirmed that Swe1p, Mps1p, Ime2p and Hrr25p readily phosphorylate poly(Tyr-Glu), but we did not detect any tyrosine kinase activity for Ste7p, Rad53p or Mck1p. Mck1p did not show strong activity in any of our assays, but Ste7p and Rad53p are very active in other assays. Thus, their inability to phosphorylate poly(Tyr-Glu) indicates that they either are very weak tyrosine kinases in general or are at least weak with the poly(Tyr-Glu) substrate. Consistent with the latter possibility, others have found that poly(Tyr-Glu) is a poor substrate for Rad53p (ref. 19; D. Stern, pers. comm.). We found that 23 other kinases also efficiently use poly(Tyr-Glu) as a substrate, indicating that there are at least 27 kinases in yeast that are capable of acting *in vitro* as tyrosine kinases. One of these, Rim11p, was recently shown to phosphorylate a Tyr residue on its *in vivo* substrate, Ime1p, indicating that it is a *bona fide* tyrosine kinase[29]. Thus, our experiment roughly tripled the number of kinases capable of phosphorylating tyrosine, and has raised questions about some of those classified as such kinases.

Correlation between functional specificity and amino sequences of the poly(Tyr-Glu) kinases

The large-scale analysis of yeast protein kinases allowed us to compare the functional relationship of the protein kinases with one another. We found that many of the kinases that phosphorylate poly(Tyr-Glu) are related to one another in their amino acid sequences: 70% of the poly(Tyr-Glu) kinases cluster into a distinct four groups on a dendrogram in which the kinases are organized relative to one another based on sequence similarity of their conserved protein kinase domains (Fig. 5a). Further examination of the amino acid sequence revealed four types of amino acids that are preferentially found in the poly(Tyr-Glu) class of kinases relative to the kinases that do not use poly(Tyr-Glu) as a substrate (three are lysines and one is a methionine); one residue (an asparagine) was preferentially located in the kinases that do not readily use poly(Tyr-Glu) as a substrate (Fig. 5b). Most of the residues lie near the catalytic portion of the molecule[30] (Fig. 5b), suggesting that they may have a role in substrate recognition.

Discussion

Large-scale analysis of protein kinases. We used a novel protein chip technology to characterize the activities of 119 protein kinases for 17 different substrates. We found that particular proteins are preferred substrates for particular protein kinases and that, vice versa, many protein kinases prefer particular substrates.
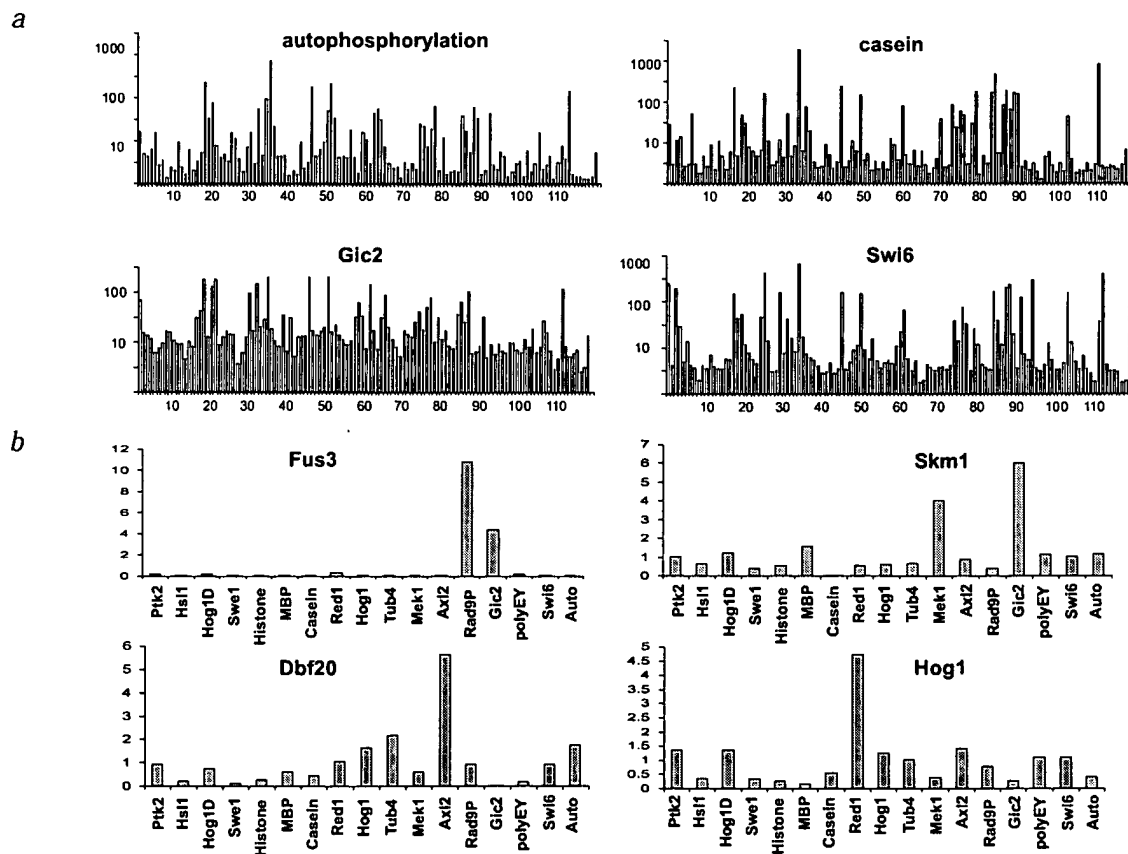
Fig. 4 Quantitative analysis of protein kinase reactions. Kinase activities were determined using a phosphoimager. The kinase signals were then transformed into fold increases by normalizing the data against negative control. *a*, Signals of 119 kinases in 4 reactions were shown in log scale. The fold increases ranged from 1 to 1000-fold. The numbers on the axis indicate the particular kinase that was analysed (for reference numbers, see Table A, http://genetics.nature.com/supplementary_info/). *b*, To determine substrate specificity, specificity index (SI) was calculated using the following formula: $SI_{ir}=F_{ir}/[(F_{i1}+F_{i2}+\cdots\cdots+F_{ir})/r]$ where *i* represents the ID of a kinase used, *r* represents the ID of a substrate, and F_{ir} represents the fold increase of a kinase *i* on substrate r compared with GST alone. Several examples of kinase specificity are shown when SI is greater than 3.The entire set of fold increase data can be retrieved from our web site (http://bioinfo.mbb.yale.edu/genome/yeast/chip).

Fig. 5 Phylogenetic tree derived from the kinase core domain multiple sequence alignment, illustrating the correlation between functional specificity and amino sequences of the poly(Tyr-Glu) kinases. **a**, Kinases that can use poly(Tyr-Glu) as a substrate often map to specific regions on a sequence comparison dendrogram. The kinases that efficiently phosphorylate poly(Tyr-Glu) are indicated in green; two kinases that weakly use this substrate are in indicated in blue. Rad53p and Ste7p, which could not phosphorylate poly(Tyr-Glu), are indicated in yellow. As shown, 70% of these kinases lie in four sequence groups (circled). **b**, Structure of the rabbit muscle phosphorylase kinase[28] (PHK). The positions of residues preferentially found in kinases that can use poly(Tyr-Glu) as a substrate are indicated in blue (dark blue indicates a basic residue; light blue indicates a methionine); the asparagine residue that is usually found in kinases that do not use poly(Tyr-Glu) is indicated in green. The conserved DFG region that is implicated in catalysis is indicated in red, whereas the conserved APE region of the substrate binding domain is indicated in purple.
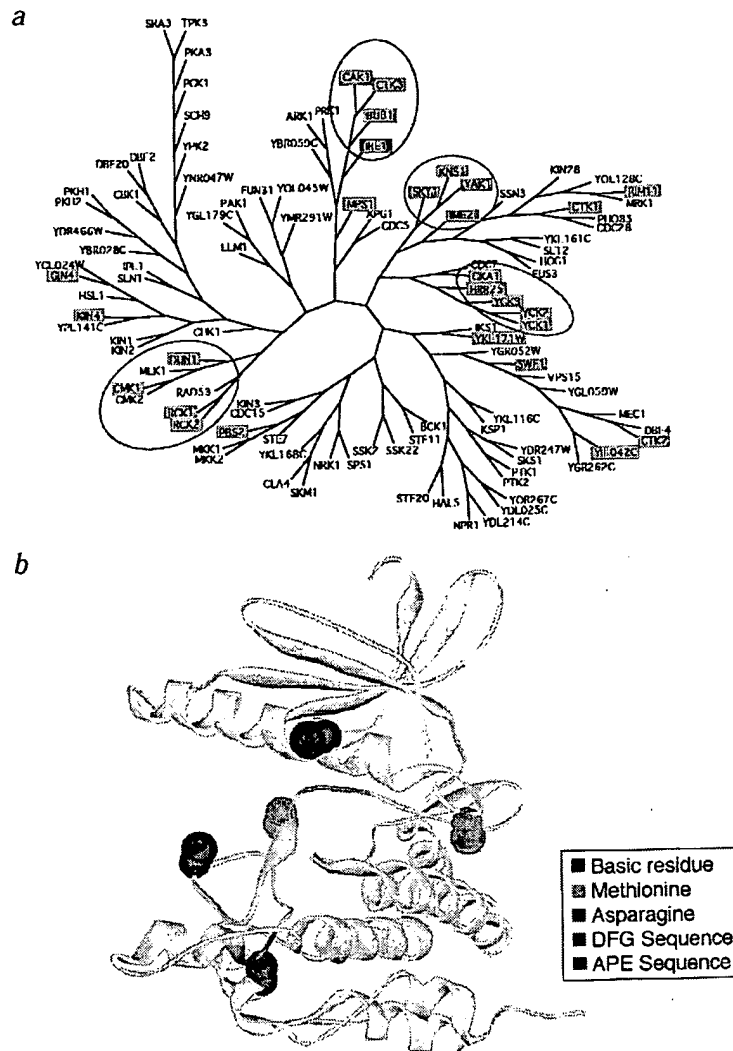
One concern with these studies is that it is possible that kinases other than the desired enzyme are contaminating our preparations. Although this cannot be rigorously ruled out, analysis of five of our samples by Coomassie staining and immunoblot staining with anti-GST antibodies does not reveal any detectable bands in our preparation that are not GST fusions.

It is important to note that *in vitro* assays do not ensure that a substrate for a particular kinase *in vitro* is phosphorylated by the same kinase *in vivo*. Other factors might restrict kinase-substrate recognition *in vivo* such as the presence of additional regulatory factors and subcellular localization. Nevertheless, these experiments indicate that certain proteins are capable of serving as substrates for specific kinases, thereby allowing further analysis. In this respect, these assays are analogous to two-hybrid studies in which candidate interactions are detected. Further experimentation is necessary to determine if the processes normally occur *in vivo*.

Consistent with the idea that many of the substrates are likely to be *bona fide* substrates *in vivo* is the observation that three kinases, Hrr25p, Pbs2p and Mek1p, phosphorylate their known substrates in our assays. Moreover, many of the kinases (for example, Ste20p) co-localize with their *in vitro* substrates (for example, Axl2p). Thus, we expect many of the kinases that phosphorylate substrates in our *in vitro* assays are likely to also do so *in vivo*.

Although most of the kinases were active in our assays, several were not. Presumably, these latter kinase preparations either lack sufficient quantities of an activator or were not purified under activating conditions. For example, Cdc28p, which was not active in our assays, might be lacking its activating cyclins. For the case of Hog1p, we treated cells with high salt to activate the enzyme. As nearly all of our kinase preparations showed activity, we presume that at least some of the enzyme in the preparation has been properly activated and/or contains the necessary cofactors. It is likely that the overexpression of these enzymes in their native organism contributes to the high success of obtaining active enzymes. It is also possible that the use of GST fusions that are capable of dimerization might augment activation of some kinases through *trans* phosphorylation. This is not the case for Hog1, which is not activated unless high salt is added to the medium.

Our assays identified many kinases that use poly(Tyr-Glu) as substrate. The large-scale analysis of many kinases allowed the novel approach of correlating functional specificity of poly(Tyr-Glu)

kinases with specific amino acid sequences. Many of the residues of the kinases that phosphorylate poly(Tyr-Glu) contain basic residues. This might be expected if there were electrostatic interactions between the kinases residues and the Glu residues. The roles of some of the other residues, however, are not obvious, such as the Met residues on the kinases that phosphorylate poly(Tyr-Glu) and the Asn on those that do not. These kinase residues may confer substrate specificity by other mechanisms. Regardless, analysis of additional substrates should allow a further correlation of functional specificity with protein kinase sequence for all protein kinases.

Protein chip technology. In addition to the rapid analysis of large number of samples, the protein chip technology described here has substantial advantages over conventional methods. First, the chip-based assays have very high signal-to-noise ratios. We found that the signal-to-noise ratio exhibited using the microwell chips is much better (>10-fold) than that observed for traditional microtitre dish assays (data not shown). Presumably this is due to the fact that [33]Pγ-ATP does not bind the PDMS as much as microtitre dishes. Second, the amount of material needed is very small. Reaction volumes are 1/20–1/40 the amount used in the 384-well microtitre dishes; less than 20 ng of protein kinase was used in each reaction. Third, the enzymatic assays using protein chips are extremely sensitive. Even though only 105 fusions were detectable by immunoblot analysis, 112 had enzymatic activity greater than fivefold over background for

at least 1 substrate. For example, Mps1p consistently exhibited the strongest activity in many of the kinase assays, even though we have never been able to detect this fusion protein by immunoblot analysis (Figs 1*b* and 3*a*). Fourth, the chips are inexpensive; the material costs less than eight cents for each array. The microfabricated molds are also easy to make and inexpensive.

In addition to the analysis of protein kinases, this protein chip technology is also applicable to a wide variety of additional assays, such as ATP and GTP binding assays, nuclease assays, helicase assays and protein-protein interaction assays. In an independent study, yeast proteins were expressed as GST fusions under the much weaker *CUP1* promotor[6]. Although the quality of these clones has not been established, biochemical activities were identified using pools of yeast strains containing the fusion proteins. The advantage of our protein chip approach is that all samples can be analysed in a single experiment. The fact that many protein kinases are active in the autophosphorylation assay indicates that at least some of the attached protein kinases retain enzymatic activity.

We used microwells that have the advantage of reducing evaporation and segregating samples, which is particularly useful for solution-based reactions. Flat PDMS chips and glass slides, however, can also be used for different assays at high density (H.Z. and M.S., unpublished data); these have the advantage that they can be used with standard pinning tool microarrayers. This technology can also be applied to facilitate high-throughput drug screening in which one can screen for compounds that inhibit or activate enzymatic activities of any gene products of interest. Because these assays will be carried out at the protein level, the results will be more direct and meaningful to the molecular function of the protein.

We configured the protein chip technology for a specific protein kinase assay using commonly available sample handling and recording equipment. For this purpose, array dimensions remained relatively large compared with dimensions readily available with micromolded silicone elastomer structures[10,31]. Thus, it should be possible to make micromolded protein chips with microwell densities increased by several orders of magnitude and carry out high-throughput biochemical assays using arrays of 10,000 to 1,000,000 microwells using automatic sample handling and measurement techniques.

We have developed an inexpensive, disposable protein chip technology for high-throughput screening of protein biochemical activity. Its usefulness was demonstrated through the analysis of 119 protein kinases from *S. cerevisiae* assayed for phosphorylation of 17 different substrates. These protein chips permit the simultaneous measurement of hundreds of protein samples. The use of micromolded microwell arrays as the basis of the chip technology allows array densities to be increased by several orders of magnitude. With the development of appropriate sample handling and measurement techniques, these protein chips may be adapted for the simultaneous assay of several thousand to millions of samples.

Methods

Cell culture, constructs and protein purification. Using a published recombination strategy[8], we cloned 119 of 122 yeast protein kinase genes in a high-copy *URA3* expression vector (pEG(KG)) that produces GST fusion proteins under the control of the galactose-inducible *GAL1* promoter[32]. Briefly, primers complementary to the end of each ORF were purchased (Research Genetics). The ends of these primers contain a common 20-bp sequence. In a second round of PCR, we modified the ends of these products by adding sequences that are homologous to the vector. The PCR products containing the vector sequences at their ends were transformed along with the vector into a *pep4* yeast strain (which lacks several yeast proteases[9]), and Ura+ colonies were selected. Plasmids were rescued into *E. coli*, verified by restriction endonuclease digestion and the DNA sequence spanning the vector-insert junction was determined using a primer complementary to the vector. For the GST::Cla4 construct, a frameshift mutation was found in a poly(A) stretch in

the amino-terminal coding region. Three independent clones were required to find the correct one that maintained reading frame. For eight kinase genes we were unable to obtain a PCR product, presumably because the genes were large. For five of these genes two overlapping PCR products were obtained and introduced into yeast cells. Confirmed plasmids were reintroduced into the *pep4* yeast strain for kinase protein purification.

For preparing samples using the 96-well format, we grew cells (0.75 ml) in medium containing raffinose to O.D.(600) ~0.5 in boxes containing 2 ml wells; two wells were used for each strain. Galactose was added to a final concentration of 4% to induce protein expression, and the cells were incubated for 4 h. The cultures of the same strain were combined, washed once with 500 μl lysis buffer, resuspended in 200 μl lysis buffer and transferred into a 96×0.5 ml plate (Dot Scientific) containing 100 μl chilled glass beads. Cells were lysed in the box by repeated vortexing at 4 °C and the GST fusion proteins were purified from these strains using glutathione beads and standard protocols[19] in a 96-well format. The purity of five purified GST::kinase proteins (Swe1, Ptk2, Pkh1, Hog1, Pbs2) was determined by comparing the Coomassie staining patterns of the purified proteins with the patterns obtained by immunoblot analysis using anti-GST antibodies. The results indicated that the purified proteins are more than 90% pure. To purify the activated form of Hog1p, cells were challenged with NaCl (0.4 M) in the last 5 min of the induction. Protein kinase activity was stable for at least 2 months at –70 °C with little or no loss of kinase activity.

Chips fabrication and protein attachment. Chips were made from the silicone elastomer PDMS (Dow Chemical) cast over micromachined molds. Liquid PDMS was poured over the molds and, after curing (at least 4 h at 65 °C), flexible silicone elastomer array sheets were peeled from the reusable molds. Although PDMS may be readily cast over microlithographically fabricated structures, for the purposes of the kinase assay described herein, molds made from sheets of acrylic patterned with a computer-controlled laser milling tool (Universal Laser Systems) sufficed.

We tested over 30 different arrays. The variables tested were width and depth of the wells (widths ranging from 100 μm to 2.5 mm, depths from 100 μm to 1 mm), spacing between wells (100 μm to 1 mm), configuration (either rectangular arrays or closest packed) and microwell shape (square versus round). The use of laser-milled acrylic molds offered a fast and inexpensive method to realize a large number of prototype molds of varying parameters.

To determine the conditions that maximize protein attachment to the wells, we treated PDMS with H_2SO_4 (5 M), NaOH (10 M), hydrogen peroxide or a crosslinker GPTS (Aldrich; ref. 11). We have found that GPTS treatment resulted in the greatest absorption of protein to the microwells relative to untreated PDMS or PDMS treated other ways. Briefly, after washing with 100% ethanol three times at RT, the chips were immersed in 1% GPST solution (95% ethanol, 16 mM HOAc) with shaking for 1 h at RT. After 3 washes with 95% ethanol, the chips were cured at 135 °C for 2 h under vacuum. Cured chips can be stored in dry argon for months[11]. To attach proteins to the chips, protein solutions were added to the wells and incubated on ice for 1–2 h. After rinsing with cold HEPES buffer (10 mM HEPES, 100 mM NaCl, pH 7.0) three times, the wells were blocked with 1% BSA in PBS (Sigma) on ice for >1 h. Because of the use of GPTS, any reagent containing primary amine groups was avoided.

To determine the concentration of proteins that can be crosslinked to the treated PDMS, HRP anti-mouse Ig (Amersham) was attached to the chip using serial dilutions of the enzyme. After extensive washing with PBS, the bound antibodies were detected using an ECL kit (Amersham). We found that up to $8×10^{-9}$ μg/μm² of protein can be attached to the surface; a minimum $8×10^{-13}$ μg/μm² is required for detection by our immunostaining methods[33].

Immunoblotting, kinase assay and data acquisition. GST::protein kinases were tested for *in vitro* kinase activity[12] using [33]Pγ-ATP. In the autophosphorylation assay, the GST:kinases were directly adhered to GPTS-treated PDMS and the *in vitro* reactions carried out with [33]Pγ-ATP in appropriate buffer. In the substrate reactions, the substrate was adhered to the wells, and the wells were washed with HEPES buffer and blocked with 1% BSA before kinase, [33]Pγ-ATP and buffer were added. The total reaction volume was kept below 0.5 μl per reaction. After incubation for 30 min at 30 °C, the chips were washed extensively, and exposed to both X-ray film and a Molecular Dynamics phosphoimager, which has a resolution of 50 μm and is quantitative. For 12 substrates each kinase assay was repeated at least twice; for the remaining 5 the assays were performed once.

Kinase sequence alignments and phylogenetic trees. Multiple sequence alignments based on the core kinase catalytic domain subsequences of the 107 protein kinases were generated with the CLUSTAL W algorithm[33], using the Gonnet 250 scoring matrix[34]. Kinase catalytic domain sequences were obtained from the SWISS-PROT (ref. 35), PIR (ref. 36) and GenBank (ref. 37) databases. For those kinases whose catalytic domains are not yet annotated (DBF4/YDR052C and SLN1/YIL147C), probable kinase subsequences were inferred from alignments with other kinase subsequences in the data set with the FASTA algorithm[38,39] using the BLOSUM 50 scoring matrix[40]. Protein subsequences corresponding to the 11 core catalytic subdomains[41] were extracted from the alignments, and the phylogenetic trees were computed with the PROTPARS (ref. 42) program (Fig. 5a).

Functional grouping of protein chip data. To visualize the approximate functional relationships between protein kinases relative to the experimental data, kinases were hierarchically ordered based on their ability to phosphorylate the 12 different substrates (data available on web site). A profile corresponding to the positive or negative activity of the 107 protein kinases to each of the substrates was recorded, with discretized values in [0,1]. Matrices were derived from the pairwise Hamming distances between experimental profiles, and unrooted phylogenies were computed using the Fitch-Margoliash least-squares estimation method[43] as implemented in the FITCH program[34] of the PHYLIP software package[42]. In each case, the input order of taxa was randomized to negate any inherent bias in the organization of the data set, and optimal hierarchies were obtained through global rearrangements of the tree structures.

1. Fields, S., Kohara, Y. & Lockhart, D.J. Functional genomics. *Proc. Natl Acad. Sci. USA* **96**, 8825–8826 (1999).
2. Goffeau, A. *et al.* Life with 6000 genes. *Science* **274**, 563–567 (1996).
3. DeRisi, J.L., Iyer, V.R. & Brown, P.O. Exploring the metabolic and genetic control of gene expression on a genomic scale. *Science* **278**, 680–686 (1997).
4. Winzeler, E.A. *et al.* Functional characterization of the *S. cerevisiae* genome by gene deletion and parallel analysis. *Science* **285**, 901–906 (1999).
5. Heyman, J.A. *et al.* Genome-scale cloning and expression of individual open reading frames using topoisomerase I-mediated ligation. *Genome Res.* **9**, 383–392 (1999).
6. Martzen, M.R. *et al.* A biochemical genomics approach for identifying genes by the activity of their products. *Science* **286**, 1153–1155 (1999).
7. Hunter, T. & Plowman, G.D. The protein kinases of budding yeast: six score and more. *Trends Biol. Sci.* **22**, 18–22 (1997).
8. Hudson, J.R. *et al.* The complete set of predicted genes from *Saccharomyces cerevisiae* in a readily usable form. *Genome Res.* **7**, 1169–1173 (1997).
9. Mitchell, D.A., Marshall, T.K. & Deschenes, R.J. Vector for the inducible overexpression of glutathione S-transferase fusion protein in yeast. *Yeast* **9**, 715–723 (1993).
10. Xia, Y. *et al.* Complex optical surfaces formed by replica molding against elastomeric masters. *Science* **273**, 347–349 (1996).
11. Rogers, Y.-H. *et al.* Immobilization of oligonucleotides onto a glass support via disulfide bonds: a method for preparation of DNA microarrays. *Anal. Biochem.* **266**, 23–30 (1999).
12. Hunter, T. & Sefton, B.M. Protein phosphorylation. *Methods Enzymol.* **200**, 35–83 (1991).
13. Roemer, T.K. *et al.* Selection of axial growth sites in yeast requires Axl2p, a novel plasma membrane glycoprotein. *Genes Dev.* **10**, 777–793 (1996).
14. Weinert, T.A. & Hartwell, L.H. Cell cycle arrest of cdc mutants and specificity of the RAD9 checkpoint. *Genetics* **134**, 63–80 (1993).
15 Jaquenoud, M., Gulli, M.P., Peter, K. & Peter, M. The Cdc42p effector Gic2p is targeted for ubiquitin-dependent degradation by the SCFGrr1 complex. *EMBO J.* **17**, 5360–5373 (1998).
16. Menees, T.M., Ross-MacDonald, P.B. & Roeder, G.S. MEI4, a meiosis-specific yeast gene required for chromosome synapsis. *Mol. Cell. Biol.* **12**, 1340–1351 (1992).
17. Bailis, J.M. & Roeder, G.S. Synaptonemal complex morphogenesis and sister-chromatid cohesion require Mek1-dependent phosphorylation of a meiotic chromosomal protein. *Genes Dev.* **12**, 3551–3563 (1998).
18. Stern, D.F., Zheng, P., Beidler, D.R. & Zerillo, C. Spk1, a new kinase from *Saccharomyces cerevisiae* phosphorylates proteins on serine, threonine, and tyrosine. *Mol. Cell. Biol.* **11**, 987–1001 (1991).
19. Kaouass, M. *et al.* The STK2 gene, which encodes a putative Ser/Thr protein kinase, is required for high-affinity spermidine transport in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **17**, 2994–3004 (1997).
20. Barral, Y., Parra, M., Bidlingmaier, S. & Snyder, M. Nim1-related kinases coordinate cell cycle progression with the organization of the peripheral cytoskeleton in yeast. *Genes Dev.* **13**, 176–187 (1999).
21. Madden, K., Sheu, Y.-J., Baetz, K., Andrews, B. & Snyder, M. SBF cell cycle regulator as a target of the yeast PKC-MAP kinase pathway. *Science* **275**, 1781–1784 (1997).
22. Sobel, S.G. & Snyder, M. A highly divergent gamma-tubulin gene is essential for cell growth and proper microtubule organization in *Saccharomyces cerevisiae*. *J. Cell. Biol.* **131**, 1775–1788 (1995).
23. Ferrigno, P., Posas, F., Koepp, D., Saito, H. & Silver, P.A. Regulated nucleo/cytoplasmic exchange of HOG1 MAPK requires the importin β homologs NMD5 and XPO1. *EMBO J.* **17**, 5606–5614 (1998).
24. Ho, U., Mason, S., Kobayashi, R., Heokstra, M. & Andrew, B. Role of the casein kinase I isoform, Hrr25, and the cell cycle-regulatory transcription factor, SBF, in the transcriptional response to DNA damage in *Saccharomyces cerevisiae*. *Proc. Natl Acad. Sci. USA* **94**, 581–586 (1997).
25. Wurgler-Murphy, S.M., Maeda, T., Witten, E.A. & Saito, H. Regulation of the *Saccharomyces cerevisiae* HOG1 mitogen-activated protein kinase by the PTP2 and PTP3 protein tyrosine phosphatases. *Mol. Cell. Biol.* **17**, 1289–1297 (1997).
26. Santos, T. & Hollingsworth, N.M. Red1p, a MEK1-dependent phosphoprotein that physically interacts with Hop1p during meiosis in yeast. *J. Biol. Chem.* **274**, 1783–1790 (1999).
27. Holly, S.P. & Blumer, K.J. PAK-family kinases regulate cell and actin polarization throughout the cell cycle of *Saccharomyces cerevisiae*. *J. Cell Biol.* **147**, 845–856 (1999).
28. Richman, T.J., Sawyer, M.M. & Johnson, D.I. The Cdc42p GTPase is involved in a G2/M morphogenetic checkpoint regulating the apical-isotropic switch and nuclear division in yeast. *J. Biol. Chem.* **274**, 16861–16870 (1999).
29. Malathi, K., Xiao, Y. & Mitchell, A.P. Catalytic roles of yeast GSK3β/shaggy homolog Rim11p in meiotic activation. *Genetics* **153**, 1145–1152 (1999).
30. Owen, D.J., Noble, M.E., Garman, E.F., Papageorgiou, A.C. & Johnson, L.N. Two structures of the catalytic domain of phosphorylase kinase: an active protein kinase complexed with substrate analogue and product. *Structure* **3**, 467–474 (1995).
31. Jackman, R.J., Duffy, D.C., Cherniavskaya, O. & Whitesides, G.M. Using elastomeric membranes as dry resists and for dry lift-off. *Langmuir* **15**, 2973–2984 (1999).
32. Mylin, L.M., Hofmann, K.J., Schultz, L.D. & Hopper, J.E. Regulated GAL4 expression cassette providing controllable and high-level output from high-copy galactose promoters in yeast. *Methods Enzymol.* **185**, 297–308 (1990).
33. Higgins, D.G., Thompson, J.D. & Gibson, T.J. Using CLUSTAL for multiple sequence alignments. *Methods Enzymol.* **266**, 383–402 (1996).
34. Gonnet, G.H., Cohen, M.A. & Benner, S.A. Exhaustive matching of the entire protein sequence database. *Science* **256**, 1443–1445 (1992).
35. Bairoch, A. & Apweiler, R. The SWISS-PROT protein sequence data bank and its supplement TrEMBL. *Nucleic Acids Res.* **27**, 49–54 (1999).
36. Barker, W.C. *et al.* The PIR-International Protein Sequence Database. *Nucleic Acids Res.* **27**, 39–43 (1999).
37. Benson, D.A. *et al.* GenBank. *Nucleic Acids Res.* **27**, 12–17 (1999).
38. Lipman, D.J. & Pearson, W.R. Rapid and sensitive protein similarity searches. *Science* **277**, 1435–1441 (1985).
39. Pearson, W.R. & Lipman, D.J. Improved tools for biological sequence comparison. *Proc. Natl Acad. Sci. USA* **85**, 2444–2448 (1988).
40. Dayhoff, M.O., Schwartz, R.M. & Orcutt, B.C. A model of evolutionary change in proteins. in *Atlas of Protein Sequence and Structure* (ed. Dayhoff, M.O.) 345–352 (National Biomedical Research Foundation, Washington DC, 1978).
41. Hanks, S.K. & Hunter, T. Protein Kinases 6. The eukaryotic protein kinase superfamily: kinase (catalytic) domain structure and classification. *FASEB J.* **9**, 576–596 (1995).
42. Felsenstein, J. PHYLIP-Phylogeny Inference Package (Version 3.2). *Cladistics* **5**, 164–166 (1989).
43. Fitch, W.M. & Margoliash, E. Construction of phylogenetic trees. *Science* **155**, 279–284 (1967).

Exhibit 20

Evidence for the denaturation of recombinant hepatitis B surface antigen on aluminium hydroxide gel

Dina Tleugabulova[a,*], Viviana Falcón[b], Eduardo Pentón[c]

[a]Quality Control Department, National Center for Bioproducts, P.O. Box 6048, Havana 6, Cuba
[b]Department of Physical Chemistry, Center for Genetic Engineering and Biotechnology, P.O. Box 6162, Havana, Cuba
[c]Vaccine Division, Center for Genetic Engineering and Biotechnology, P.O. Box 6162, Havana, Cuba

Abstract

Despite the complexity of the subject of protein–alum interactions, a valuable information can be obtained by analyzing the adsorbed and desorbed protein by common physico–chemical methods. In the present work, to approach the adsorption of hepatitis B surface antigen (HBsAg) on alum, the experimental data were supported by complementary analyses of the adsorbed protein by immunoelectron microscopy and the desorbed protein by denaturing size-exclusion chromatography and sodium dodecyl sulfate-polyacrylamide gel electrophoresis under reducing conditions. First, the depletion of HBsAg was investigated. The aspects assessed were the conditions, recovery and chromatographic performance of the desorbed protein. The results obtained strongly suggested the loss of particulate structure of HBsAg after adsorption on alum. This conclusion was further reinforced by direct immunoelectron microscopic visualization of HBsAg in the adsorbed state.   © 1998 Published by Elsevier Science B.V. All rights reserved.

1. Introduction

According to World Health Organization estimates, by the year 2000, there will be 400 million hepatitis B virus carriers in the world if hepatitis B vaccine is not widely used [1]. Several yeast-derived hepatitis B vaccines are commercially available now based on the same recombinant hepatitis B surface antigen (HBsAg) and aluminium hydroxide (alum) adjuvant. In numerous clinical trials, these preparations have demonstrated an immunogenicity and efficacy similar to that of plasma-derived antigen, with no difference in antibody specificity and avidity

*Corresponding author.

[2–4]. However, using both plasma- and yeast-derived vaccines, a small portion (about 5–10%) of vaccinated subjects fail to produce detectable antibody response. Induction of immune reactivity depends upon antigen reaching and being available in lymphoid organs in a dose- and time-dependent manner [5]. Since HBsAg is administered as the alum-adsorbed preparation, antigen–alum interactions should be crucial in the immune response. This was experimentally shown for gp120 protein [6]. As most proteins, gp120 is quantitatively adsorbed on alum in a few minutes. However, the gp120–alum interaction is weak and sensitive to anions from physiological fluids. Due to this, the protein is rapidly desorbed from alum just after injection that

explains a weak immune response to the MN gp120 HIV-1 vaccine.

Our understanding of antigen–alum interactions may provide a way to understand the non-responsiveness of some people to hepatitis B vaccination. The analysis of antigen–alum interactions requires knowledge of both the kinetic events that occur during the adsorption event and the structure of adsorbed protein. However, neither of these aspects has been explored for alum-adsorbed proteins due to the lack of reliable experimental techniques capable to support kinetic measurements as well as to assess the adsorbed protein at a molecular level [Review, 7]. Despite the complexity of alum structure [8] and great structural variety among antigens, protein adsorption on alum is commonly approached by Langmuir model originally developed for small molecules adsorbing at an 'ideal' surface. This model cannot describe the adsorption behavior of several antigen proteins [9], including HBsAg [unpublished results]. However, without the stimulus of usable data, no superior theoretical treatment has been developed.

Despite the complexity of the subject of protein–alum interactions, a valuable information can be obtained by analyzing the adsorbed and desorbed protein by common physico–chemical methods. In the present work, to approach the adsorption of hepatitis B surface antigen (HBsAg) on alum, the experimental data were supported by complementary analyses of the adsorbed protein by immunoelectron microscopy and the desorbed protein by denaturing size exclusion chromatography (SEC) and sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE). The results obtained evidence the loss of particulate structure of HBsAg after adsorption on alum. The same has been previously suggested from the immunological studies of HBsAg [10] and another virus-like particles [11,12]. Hence, our work provides a physico–chemical support to this hypothesis.

2. Experimental

2.1. Materials

Tris(hydroxymethyl)aminomethane, dithiotreitol (DTT), SDS, mercaptoethanol, sodium chloride,

sodium phosphates and another mentioned reagents were analytical grade and obtained from Merck (Darmstadt, Germany). The reagents used in electron microscopy were from Agar Scientific (Essex, UK). All solutions were made in Milli-Q grade water. Aluminium hydroxide gel (alum) was purchased as a sterile 2% (w/v) Al(OH)$_3$ suspension from Superfos Biosector (Vedbaek, Denmark). Hyflo Super Cel (celite) was from Fluka (Buchs, Switzerland). Phosphate-buffered saline (PBS) contained 1.7 mM KH$_2$PO$_4$, 7.9 mM Na$_2$HPO$_4$, 2.7 mM KCl and 250 mM NaCl, pH 7.0. Recombinant HBsAg, cloned and expressed in yeast Pichia pastoris, was purified by a multi-step procedure [13] and provided as a solution in PBS (1.61 mg/ml) from the National Center for Bioproducts (Havana, Cuba). This stock solution was used in the preparation of working standard solutions of lower concentrations as well as in the adsorption studies.

Anti-HBsAg mouse monoclonal antibody (CB Hep1) and protein A–colloidal gold complex (particle diameter, 15 nm) used for the immunodetection of HBsAg were provided by the Division of Immunotechnology and Diagnostics of the Center for Genetic Engineering and Biotechnology (Havana, Cuba). ~

The HBsAg–celite preparation (100 μg/ml HBsAg) used as a reference standard in the electron microscopic study of HBsAg–alum preparation was prepared as described in Ref. [14].

2.2. Apparatus

2.2.1. Size-exclusion chromatography (SEC)

The SEC system included a Pharmacia LKB 2248 pump, Knauer degasser, Pharmacia 2141 variable-wavelength UV detector operated at 280 nm and Pharmacia 2221 programmable integrator. The column used was a TSK G4000 SW (600×7.5 mm I. D.) purchased from Tosohaas (Stuttgart, Germany). Elution was with 0.1 M Tris–HCl in 0.3% SDS, pH 8.0 at 0.9 ml/min. After injecting the working standard solutions (100 μl), the conversion of peak areas to protein concentrations was carried out using programmable integration.

2.2.2. SDS-PAGE

Electrophoresis (Hoefer Scientific Instrument) was performed as described by Laemmli [15] on 12.5%

gels at 30 mA for 3.5 h at room temperature under reducing conditions. The gels were stained by Coomasie blue dye (Bio-Rad, Richmond, CA, USA) or silver nitrate [16]. The Coomasie-stained gel was scanned by laser densitometry using Ultroscan XL (Pharmacia). For immunoblotting, the gel was incubated with CB Hep1 antibody and developed with protein A conjugated to aminobenzidine [17].

2.3. Preparation of HBsAg–alum

A mixture of 2% aluminium hydroxide gel (2 ml) and HBsAg stock solution (1.7 ml) was gently agitated for 30 min at room temperature and then diluted with PBS in a 25-ml volumetric flask. The concentration of adsorbed HBsAg was 100 μg/ml, as determined by Lowry method [18] from protein balance. The placebo was prepared by the dilution of 2% aluminium hydroxide gel (2 ml) with PBS up to 25 ml.

2.4. Reduction of HBsAg particles

Aliquots (200 μl) from working standard solutions (0.2–1.6 mg/ml HBsAg) were incubated with DTT/ M sample buffer (40 μl) for 10 min at 100°C [The DTT/M sample buffer contained 417 mM DTT, 4.2% (w/v) SDS and 16% 2-mercaptoethanol]. The reduced samples were analyzed by SDS-PAGE/ Coomasie blue staining (30 μl) and SEC (100 μl) as described.

2.5. Desorption of HBsAg from alum

One ml of the HBsAg–alum preparation (100 μg/ml) was centrifuged for 10 min at 2500 rpm. The pellet separated was incubated for 3 min at 100°C with a mixture of 0.4 M Na/PO$_4$, pH 8.0 (100 μl)–DTT/M sample buffer (20 μl) (5:1). After centrifugation for 5 min at 10 000 rpm, the supernatant was analyzed by SDS-PAGE/Coomasie blue staining (30 μl) and SEC (100 μl).

In order to determine the HBsAg recovery after desorption, the area of SEC-peak from the desorbed sample was extrapolated to the calibration curve previously generated by injections of the HBsAg working standard solutions. The determinations were carried out in duplicate.

In another experiment (Fig. 5), the HBsAg was

desorbed from alum using the described procedure, except the incubations at 100°C were for 10, 15, 20 and 25 min, respectively. After that, the samples were analyzed by SEC (100 μl) and SDS-PAGE/ silver staining (1.7 μg HBsAg per spot).

2.6. Transmission electron microscopy (TEM)

Two drops of HBsAg solution in PBS (0.1 mg/ ml) were placed for 5 min onto a 400 mesh copper grid coated with formvar-carbon film. Excess sample was blotted off. Grids were stained with uranyl acetate and examined in a Jeol-JEM 2000EX transmission electron microscope, acceleration voltage 80 kV and magnification 40 000X.

The adsorbed HBsAg was analyzed by sectioning the HBsAg–alum and HBsAg–celite pellets obtained by centrifugation for 10 min at 2500 rpm of the respective preparations (10 ml). Placebo was used as a blank. The pellets were fixed by immersion in 1% glutaraldehyde, rinsed with PBS and dehydrated in increasing (30–100%) ethanol concentrations. The embedding was in Araldite. Each resulting block was sectioned (N=60) with an ultramicrotome LKB 2188 (NOVA) and 400-Å sections were mounted on 400 mesh nickel grids without membrane. Staining was with uranyl acetate and lead citrate followed by examination as described.

2.7. Immunoelectron microscopy

The grids coated with soluble HBsAg were floated on six drops of gold buffer (1% BSA in PBS) before transfer to a drop of the PBS-diluted CB Hep1 antibody (dilution, 1:20) for incubation at room temperature (30 min). After washing in gold buffer to remove unbound antibody molecules, the grids were floated on two drops of gold buffer-diluted protein A–gold complexes (dilution, 1:200) for 40 min at room temperature. Finally, the grids were subsequently washed in six drops of gold buffer and two drops of distilled water. After staining with 1% uranyl acetate, the grids were examined as described.

Similarly, the sections (N=12) from the placebo, HBsAg–alum and HBsAg–celite pellets were incubated for 5 min on a drop of PBS containing 1% ovalbumin before transfer to a drop of PBS-diluted CB Hep1 monoclonal antibody (dilution 1:20). Incubation was for 60 min at room temperature. The

grids were then rinsed for 5 min with PBS to remove unbound antibody molecules and then incubated on a drop of protein A–colloidal gold complex, pH 7.2, for 40 min at room temperature. Finally, the sections were stained and examined as described.

3. Results and discussion

3.1. Analysis of reduced HBsAg

Recombinant HBsAg is produced by the expression of a 226-amino acid polypeptide in yeast cells where approximately 100 of these polypeptides are assembled intracellularly into 22 nm lipoprotein particles [19]. After purification, the assembled HBsAg particles are detected by SEC, electron microscopy and enzyme-linked immunosorbent assay (ELISA) using polyclonal antibodies [20]. To assess the HBsAg monomer, the particles should be previously reduced. Efficient reduction of HBsAg has been achieved by using a mixture of DTT and mercaptoethanol (DTT/M), instead of one of them alone [21,22], suggesting that these reducing agents do not possess the same reducing power on multiple disulfide bonds within HBsAg particles. Soluble HBsAg–SDS complexes formed after reduction are suitable for further analysis by reversed-phase high-performance liquid chromatography (RP-HPLC) [21] or denaturing SEC [22].

In denaturing SEC (Fig. 1), the reduced HBsAg is resolved into the three peaks: peak 1 corresponds to the co-elution of non-reduced HBsAg and non-protein micellar aggregates, peak 2 to the HBsAg dimers and monomers and peak 3 to lower-molecular-mass non-protein compounds. Earlier studies have shown that peak 1 remaining after DTT/M reduction is produced by non-protein aggregates, probably, by lipid–SDS complexes [22].

The correlation between the areas of peak 2 and the HBsAg concentrations taken for reduction was linear in the range 0.2–1.6 mg/ml HBsAg ($r=$ 0.9991). The reproducibility was tested with three replicate injections of the reduced stock solution of HBsAg on three different days. The relative standard deviations (R.S.D.s) were 1.16–1.36% (within-days) and 1.25–1.65% (between-days).

The presence of shoulders before and after the
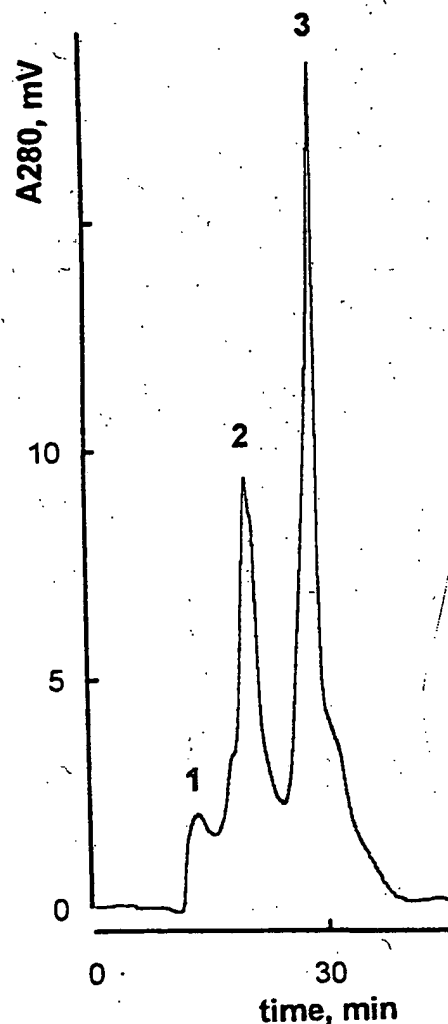
Fig. 1. Chromatogram of reduced HBsAg. Conditions: TSK G4000 SW (600×7.5 mm I.D.); eluent, 0.1 M Tris–HCl containing 0.3% SDS, pH 8.0; flow rate, 0.9 ml/min; detection, UV at 280 nm; injection volume, 100 μl; sample buffer, 417 mM DTT, 4.2% (w/v) SDS and 16% (v/v) 2-mercaptoethanol. Peaks: 1= Lipid–SDS aggregates, 2=reduced HBsAg, 3=low-molecular-mass non-protein compounds.

maximum of peak 2 (Fig. 1) was a hint at the heterogeneity of HBsAg structures formed after reduction. Indeed, as shown by SDS-PAGE/silver staining of SEC-fractions from peak 2 (Fig. 2), the reduced HBsAg was represented by dimers eluting presumably in the forward shoulder, monomers eluting in the maximum of peak 2 and lower-molecular-mass proteins eluting in the backward shoulder.
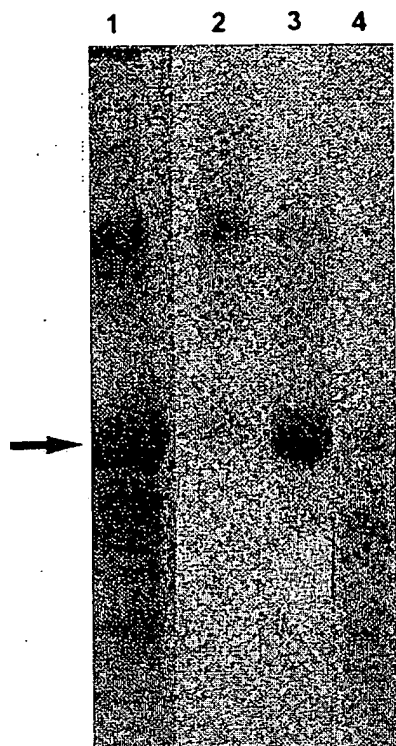
Fig. 2. SDS-PAGE/silver staining of DTT/M-reduced HBsAg (lane 1) subjected to SEC fractionation (lanes 2–4): forward shoulder of peak 2 (lane 2), maximum of peak 2 (lane 3) and backward shoulder of peak 2 (lane 4). Experimental conditions as in text. The arrow indicates the M_r 24 000 HBsAg monomer. Amount, 20 (lane 1) and 1.7 (lanes 2–4) μg HBsAg.

The last ones migrated before the M_r 24 000 monomer band on SDS-gel (Fig. 2, lane 4) and were recognized in immunoblotting (data not shown). When HBsAg monomer from the maximum of peak 2 (Fig. 2, lane 3) was repeatedly reduced, no appearance of lower bands was detected evidencing that the observed lower bands are not generated by the reducing procedure. We assumed these bands as degradation products. Similar degradation bands were detected by immunoblotting of HBsAg expressed in *S. cerevisiae* or *H. polymorpha* [23]. Since these bands were found not only in purified material but also in crude HBsAg-containing yeast extract [23], the detected degradation probably takes place in vivo. The degradation products are particle-associated, because we could not separate them from assembled particles by SEC under non-denaturing conditions [data not shown]. Since HBsAg particles

are resistant to proteases [24], the degradation probably implies covalent modifications of labile amino acids exposed on particle surface (Ser, Thr β-elimination and racemization, Asn deamidation, Cys, Trp, Tyr oxidation or the hydrolysis of peptide bonds [25]).

3.2. Analysis of desorbed HBsAg

To improve the immunogenicity of purified HBsAg particles, these are adsorbed onto aluminium hydroxide gel in hepatitis B vaccine. Like tetanus toxoid and diphtheria toxoid [9], HBsAg is adsorbed on alum independently on pH and excess phosphate ions, and this adsorption is irreversible under non-denaturing conditions [unpublished results]. In the present work, adsorbed HBsAg was recovered after the reduction of HBsAg–alum pellet with a mixture of 0.4 M Na/PO$_4$, pH 8.0–DTT/M sample buffer (Fig. 3). The desorbed protein migrated presumably as a M_r 24 000 monomer onto SDS-gel. When the known amounts of HBsAg were reduced and applied on gel, the intensities of the M_r 24 000-bands detected by laser densitometry were linearly related to the HBsAg amounts ($r=0.995$). By extrapolating the intensity of the band from the desorbed sample, the HBsAg recovery was estimated to be 43±4%. This value may be overestimated due to a significant enlargement of the band from desorbed HBsAg compared to those from HBsAg solutions (Fig. 3). As an alternative, the desorbed HBsAg was determined by SEC. The chromatogram of desorbed HBsAg was quite similar to that of HBsAg reduced in solution (Fig. 4). By plotting the area of peak 2 from the desorbed sample on the calibration curve, the HBsAg recovery was calculated to be 50±1%. Hence, both SDS-PAGE and SEC indicate that a large portion of adsorbed HBsAg cannot be recovered under reducing conditions.

In respect to the recoverable fraction of HBsAg, it was continuously degraded at increasing the rates of reduction in the range from 3 to 25 min. In the chromatogram of desorbed HBsAg, the height of peak 2 was gradually diminished, whereas the backward shoulder corresponding to the elution of degraded polypeptides increased. After 20 min of boiling, the desorbed HBsAg was presented mainly as the protein fragments. Under the same conditions,
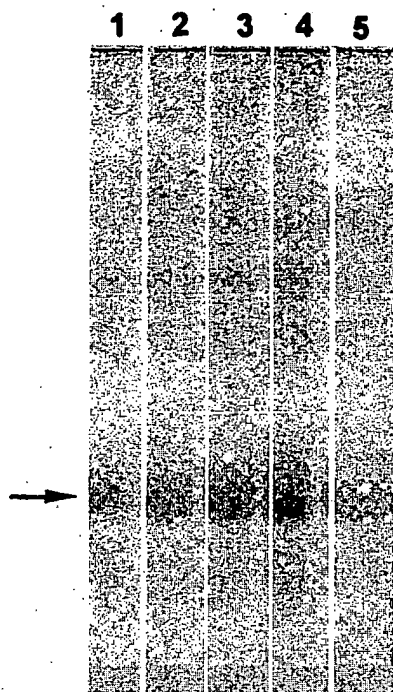
Fig. 3. SDS-PAGE/Coomasie blue staining of HBsAg standard solutions (lanes 1–4) and alum-desorbed sample. Experimental conditions as in Section 2. The arrow indicates the M_r 24 000 HBsAg monomer. Amount, 15 (lane 1), 20 (lane 2), 25 (lane 3) and 30 (lane 4) µg HBsAg. Desorbed sample: the value expected assuming the complete recovery of HBsAg, 20 µg; the value calculated from the calibration curve, 9 µg.

Fig. 4. Chromatogram of HBsAg reduced in solution (A) and desorbed from alum under reducing conditions (B). Conditions as in Fig. 1; concentration, 0.5 mg/ml HBsAg (A); injection volume, 100 µl. Desorbed HBsAg: the value expected, assuming complete HBsAg recovery, 0.5 mg/ml HBsAg; the value calculated from the calibration curve, 0.24 mg/ml HBsAg.

the HBsAg monomer from intact particles was stable and no changes in its chromatographic profile (Fig. 1) were observed even after 30 min of boiling with DTT/M sample buffer. The degradation detected by SEC for the recoverable fraction of adsorbed HBsAg was almost undetectable by SDS-PAGE, where larger rates of boiling led to a weak increase in the intensity of degradation bands (Fig. 5, right). This is probably due to a highly intensive, selective staining of HBsAg bands compared to that of degradation bands.

In conclusion, the present study of HBsAg desorption raised two important observations. First, only half protein adsorbed is recovered under reducing conditions. Since the reducing conditions are capable to efficiently disrupt any type of interactions having these an ionic, hydrophobic and/or ligand-exchange nature, the non-recoverable fraction of HBsAg

should be trapped to alum by an additional, besides adsorption, mechanism making protein molecules inaccessible to reducing buffer. If the particulate structure of HBsAg were to be preserved in the adsorbed state, the HBsAg monomers would be quantitatively recovered. Second, the fraction of HBsAg recoverable under reducing conditions is prone to temperature-induced degradation, unlike intact HBsAg. Since the stability of a protein is determined by its three-dimensional structure [25], the results indicate that the structure of HBsAg particles is altered by alum adjuvant.

The increasing attention to particulate polymeric antigens in the form of virus-like particles is related to their ability to induce specific, cell-mediated immune response in the absence of adjuvants [Reviews, 26, 27]. Cytotoxic T lymphocytes (CTLs) provide a critical arm of the immune system in eliminating autologous cells expressing foreign antigen [28–30]. Although the mechanisms by which these approaches lead to the induction of CTLs are
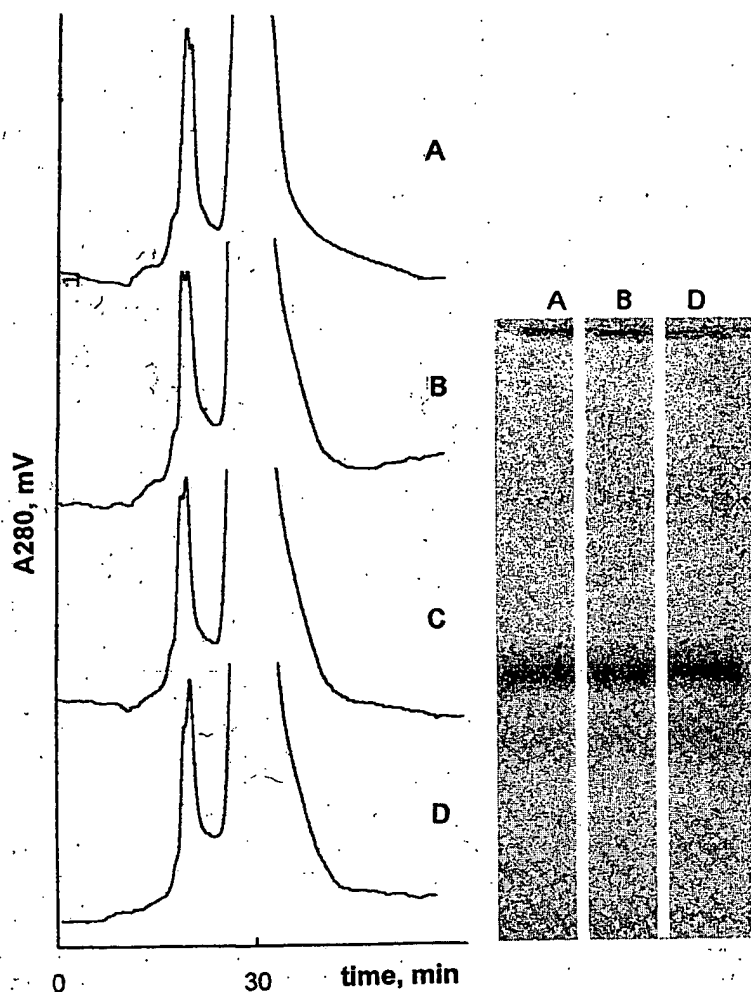
Fig. 5. Chromatographic profile (left) and SDS-PAGE/silver staining pattern (right) of desorbed HBsAg after 10 (A), 15 (B), 20 (C) and 25 (D) min of boiling the HBsAg–alum pellet with DTT/M buffer. Chromatographic conditions as in Fig. 1; injection volume, 100 μl. Electrophoretic conditions as in Section 2; amount, 1.7 μg HBsAg.

unknown, it has been suggested that the particulate nature of virus-like particles favors their optimal delivery to the class I antigen presentation pathway [11,12,31]. Unexpectedly, particulate antigens, including HBsAg, failed to stimulate CTLs after their adsorption on alum eliciting, in contrast, substantial antibody titers [10, 11]. Hence, a supposition has been made that the particulate structure of virus-like particles may be compromised by the adsorption on adjuvant [31]. Our results from SEC and SDS-PAGE support this feeling. In order to provide a direct evidence for the loss of particulate structure of

HBsAg on alum, we analyzed the HBsAg–alum preparation by immunoelectron microscopy.

3.3. Immunoelectron visualization of adsorbed HBsAg

The intact HBsAg was seen in electron microscope as 22-nm spheres (Fig. 6A). After its adsorption on alum, no particles were found in the HBsAg–alum sections suggesting alterations in the HBsAg morphology (Fig. 6B). In a few sections, a structures were found resembling HBsAg particles,
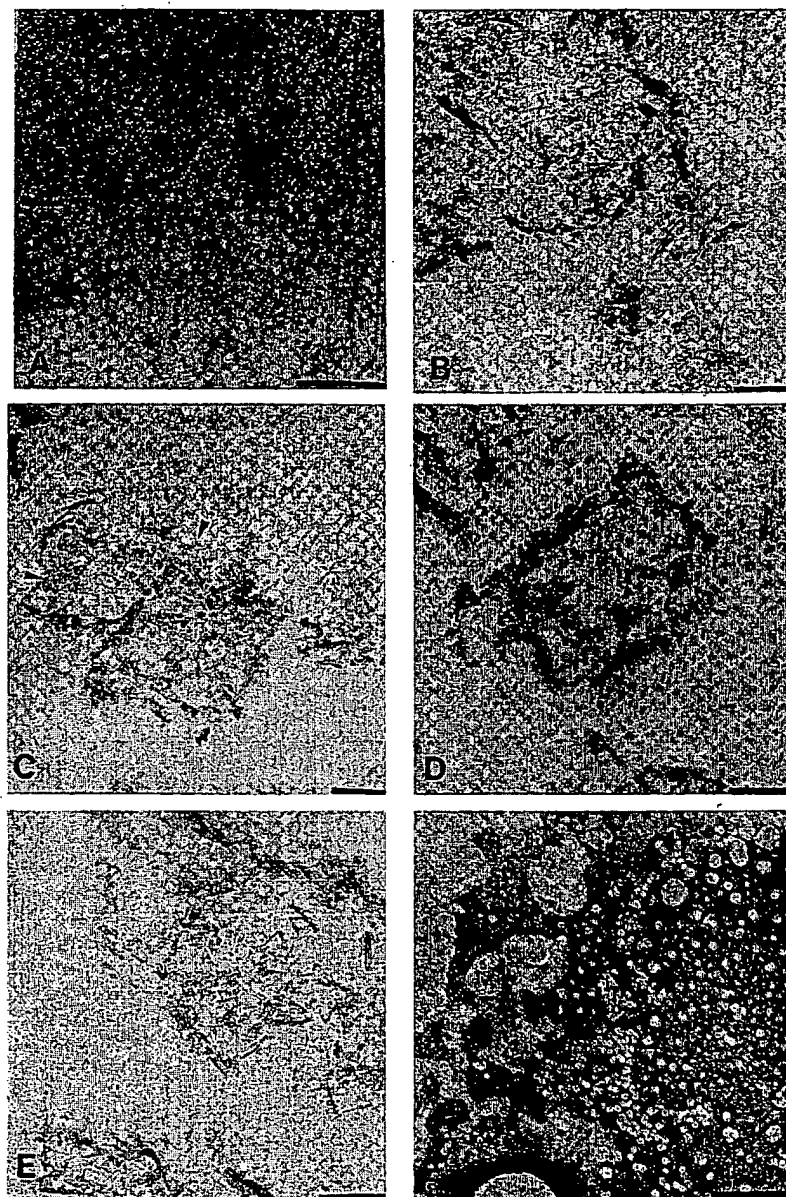
Fig. 6. Electron micrographs of intact HBsAg (A), HBsAg–alum sections (B, C and D), placebo (E) and HBsAg–alum (F) sections. Experimental conditions as in Section 2. Scale bar is 200 nm.

but with a defect in particulation, suggesting particle damage (Fig. 6C and D). These structures were absent in the placebo sections (Fig. 6E). In order to prove that HBsAg structure is unaffected under the conditions of analysis, the HBsAg was adsorbed on celite and the HBsAg–celite sections were prepared

as described. The HBsAg adsorption on celite is reversible [14], thus, the HBsAg particles should be found in the HBsAg–celite sections. As expected, the HBsAg particles retained in the pores of celite were clearly visualized (Fig. 6F). The adsorbed HBsAg, undetectable by direct visualization, was

identified here by immunoelectron microscopy (IEM) using CB Hep1 monoclonal antibody. One of the most interesting aspects of this approach is its potential to detect HBsAg independently on the state of folding. When the HBsAg–alum sections were incubated with CB-Hep 1 antibody followed by successive labeling with protein A–gold complex, these were specifically labeled proving thus the presence of HBsAg on alum gel (Fig. 7A and B). The HBsAg appeared as in a dark-staining matrix structure of Al(OH)$_3$ (Fig. 7A) as in the gel pores (Fig. 7B). It is essential that non-specific adsorption of gold on alum was minimal in the absence of HBsAg (Fig. 7C).

In IEM of intact HBsAg, the mapping of 15-nm gold onto 22-nm HBsAg particles produces a labeling pattern characterized by the presence of hoops around gold particles (Fig. 7D). This pattern was not observed after immunolabeling of HBsAg–alum sections (Fig. 7A and B) proving once more the loss of particulate structure of HBsAg after adsorption on alum.

3.4. Proposed model

The initial adsorption event is a function of the interfacial activity of various groups on a particle surface and on the adsorbent [32]. The active sites on alum are ascribed to aluminium ions (Lewis acids) or hydroxyls [7], whereas those on HBsAg particle are represented by water-exposed phospholipids and hydrophilic protein domains [33]. The HBsAg–alum interaction is of non-ionic character, because it is insensitive to the conditions of pH and ionic strength. Despite ionic interaction, proteins are adsorbed on alum by hydrophobic [34] and ligand-exchange [35,36] forces. The last mechanism comprises the chemisorptive binding of phosphate and carboxylate groups of a protein to alum surface in accord with the Lewis acid–base model. Taking into
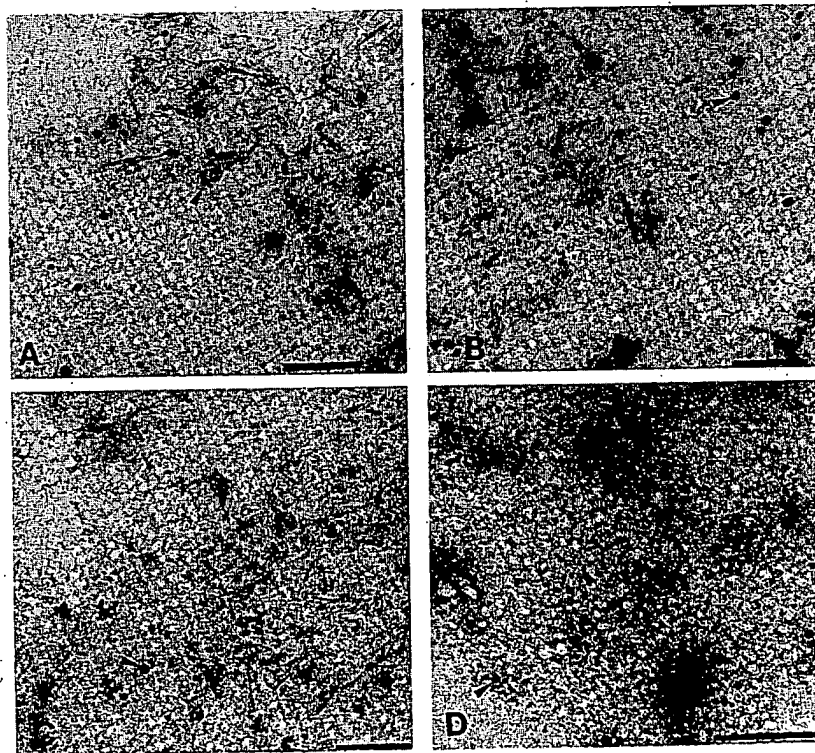
Fig. 7. Protein A–gold electron microscopic localization of HBsAg on alum (A, B). Non-specific labeling of alum in the absence of HBsAg (C). Immunolabeling pattern of intact HBsAg (D). Experimental conditions as in Section 2. Scale bar is 200 nm.

account that phosphate is one of the strongest Lewis bases, the interaction of phosphate heads from phospholipids of HBsAg with unsaturated aluminium ions is highly expected. Phospholipids are the pre-dominant lipid structures within HBsAg particles [37] and play a crucial role in the particle stabiliza-tion [38]. The lipid–protein interactions are respon-sible for the formation of the proper helical structure of the HBsAg protein, which disposes the remainder of the protein on the surface or interior of the particle. The lipid–protein interactions stabilize the conformation of the exterior hydrophilic regions which contain the HBsAg antigenic sites [38]. Hence, the adsorption of phospholipids on alum is expected to produce a damage of integrity of outer lipid monolayer followed by rearrangements in hy-drophobic protein domains and lipid core inside the HBsAg particle. According to this model, the ob-served low recovery of HBsAg under reducing conditions can be explained by a trap of the non-recoverable protein monomers between alum and disordered lipids from lipid core. On the other hand, the alterations in lipid–protein interactions within HBsAg particle destabilize α-helices by the exposure of protein domains previously buried in lipid bilayer [38]. Being less stable, the conformationally altered HBsAg monomers are more sensitive to degradation that explains the observed temperature-induced deg-radation of the recoverable HBsAg monomers.

Acknowledgements

References

[1] from Hepatitis Weekly, July 8 (1996) 1.
[2] W. Jilg, F. Deinhardt, J. Infect. 13 (1986) 47.
[3] W. Jilg, B. Lorbeer, M. Schmidt, B. Wilske, G. Zoulek, F. Deinhardt, Lancet 11 (1984) 1174.
[4] F. Lugoboni, S. Migliozzi, F. Schiesari, N. Pauletto, G.L. Bovo, S. Ciaffoni, P. Mezzelani, Vaccine 15 (1997) 1014.
[5] R.M. Zinkernagel, S. Ehl, P. Aichele, S. Oehen, T. Kundig, H. Hengartner, Immunol. Rev. 156 (1997) 199.
[6] R.P. Weissburg, P.W. Berman, J.L. Cleland, D. Eastman, F. Farina, S. Frie, A. Lim, J. Mordenti, M.R. Peterson, K. Yim, M.F. Powell, Pharm. Res. 12 (1995) 1439.
[7] S.L. Hem, J.L. White, in: M.F. Powell, M.J. Newman (Eds.), Vaccine Design: The Subunit and Adjuvant Approach, Plenum Press, New York, 1995, p. 249.
[8] S.L. Hem, J.L. White, J. Parenteral Sci. Tech. 38 (1984) 2.
[9] R.K. Gupta, B.E. Rost, E.H. Relyveld, G.R. Siber, in: M.F. Powell, M.J. Newman (Eds.), Vaccine Design: the Subunit and Adjuvant Approach, Plenum Press, New York, 1995, p. 229.
[10] R. Schirmbeck, K. Melber, T. Mertens, J. Reimann, Eur. J. Immunol. 24 (1994) 1088.
[11] G.T. Layton, S.J. Harris, A.J. H. Gearing, M. Hill-Perkins, J.S. Cole, J.C. Griffiths, N.R. Burns, A.J. Kingsman, S.E. Adams, J. Immunol. 151 (1993) 1097.
[12] S.J. Harris, S.A. Woodrow, A.J. H. Gearing, S.E. Adams, A.J. Kingsman, G.T. Layton, Vaccine 14 (1996) 971.
[13] E. Pentón, Eur. Pat. Publ. No. EP 480525 (1992).
[14] A. Agráz, Y. Quiñones, N. Expósito, F. Breña, J. Madruga, E. Pentón, L. Herrera, Biotechnol. Bioeng. 42 (1993) 1238.
[15] U.K. Laemmli, Nature 227 (1970) 680.
[16] W. Wray, T. Boulikas, V.P. Wray, R. Hancock, Anal. Bio-chem. 118 (1981) 197.
[17] O. Rodriguez, M. Izquierdo, Y. Martinez, A. Garcia, Adv. Biotecnol. Moderna 2 (1994) 64.
[18] O.H. Lowry, N.J. Rosebrough, A.L. Farr, R.J. Randall, J. Biol. Chem. 265 (1951) 265.
[19] P. Valenzuela, A. Medina, W.J. Rutter, G. Ammerer, B.D. Hall, Nature 298 (1982) 347.
[20] D. Tleugabulova, J. Reyes, L. Costa, J. Díaz, J.J. Madrazo-Piñol, Chromatographia 45 (1997) 317.
[21] D.O. O'Keefe, A.M. Paiva, Anal. Biochem. 230 (1995) 48.
[22] D. Tleugabulova, J. Chromatogr. B 713 (1998) 401.
[23] Z.A. Janowicz, K. Melber, A. Merckelbach, E. Jacobs, N. Harford, M. Comberbach, C.P. Hollenberg, Yeast 7 (1991) 431.
[24] R. Prange, R.E. Streeck, EMBO J. 14 (1995) 247.
[25] S. Li, Ch. Schoneich, R. Borchardt, Biotech. Bioeng. 48 (1995) 490.
[26] S.C. Gilbert, M. Plebanski, S.J. Harris, C.E.M. Allsopp, R. Thomas, G.T. Layton, A.V.S. Hill, Nature Biotechnol. 15 (1997) 1280.
[27] P.D. Robbins, H. Tahara, S.C. Ghivizzani, Trends Biotech-nol. 16 (1998) 35.
[28] Y.L. Lin, B.A. Askonas, J. Exp. Med. 154 (1981) 225.
[29] M.M. McNeal, K.S. Barone, M.N. Rae, R.L. Ward, Virology 214 (1995) 387.
[30] P. Reusser, S.R. Riddell, J.D. Meyers, P.D. Greenberg, Blood 78 (1991) 1373.
[31] C. Sedlik, M.F. Saron, J. Sarraseca, I. Casal, C. Leclerc, Proc. Natl. Acad. Sci. USA 94 (1997) 7503.
[32] J.D. Andrade, V. Hlady, A.P. Wei, C.G. Golander, Croatica Chem. Acta 63 (1990) 527.
[33] M. Sato, Y. Sato-Miyamoto, K. Kameyama, N. Ishikawa, M. Imai, Y. Ito, T. Takagi, J. Biochem. 118 (1995) 1297.

[34] A. Bhaduri, N. Matsudomi, K.P. Das, Biosci. Biotech. Biochem. 60 (1996) 1559.

[35] R.J. Sepelyak, J.R. Feldkamp, T.E. Moody, J.L. White, S.L. Hem, J. Pharm. Sci. 73 (1984) 1514.

[36] M.J. Gorbunoff, Anal. Biochem. 136 (1984) 425.

[37] F. Gavilanes, J.M. Gonzalez-Ros, D. Peterson, J. Biol. Chem. 257 (1982) 7770.

[38] F. Gavilanes, J. Gomez-Gutierrez, M. Aracil, J.M. Gonzalez-Ros, J.A. Ferragut, E. Guerrero, D. Peterson, Biochem. J. 265 (1990) 857.

JOURNAL OF CHROMATOGRAPHY B

BIOMEDICAL SCIENCES AND APPLICATIONS

ELSEVIER

J|A|C|S

Exhibit 21

A R T I C L E S

Biofunctionalized, Ultrathin Coatings of Cross-Linked Star-Shaped Poly(ethylene oxide) Allow Reversible Folding of Immobilized Proteins

Jürgen Groll,[†] Elza V. Amirgoulova,[‡] Thomas Ameringer,[†] Colin D. Heyes,[‡]
Carlheinz Röcker,[‡] G. Ulrich Nienhaus,*[‡,§] and Martin Möller*[||]

*Contribution from the Department of Macromolecular Chemistry and Organic Chemistry III,
University of Ulm, Albert Einstein Allee 11, 89081 Ulm, Germany, Department of Biophysics,
University of Ulm, Albert Einstein Allee 11, 89081 Ulm, Germany, Department of Physics,
University of Illinois at Urbana-Champaign, Urbana, Illinois, 61801, and Deutsches
Wollforschungsinstitut an der RWTH Aachen und Institut für Technische und Makromolekulare
Chemie der RWTH Aachen, Veltmanplatz 8, 52062 Aachen, Germany*

Received December 18, 2003; E-mail: moeller@dwi.rwth-aachen.de; uli@uiuc.edu.

Abstract: Dense, ultrathin networks of isocyanate terminated star-shaped poly(ethylene oxide) (PEO) molecules, cross-linked at their chain ends via urea groups, were shown to be extremely resistant to unspecific adsorption of proteins while at the same time suitable for easy biocompatible modification. Application by spin coating offers a simple procedure for the preparation of minimally interacting surfaces that are functionalized by suitable linker groups to immobilize proteins in their native conformations. These coatings form a versatile basis for biofunctional and biomimetic surfaces. We have demonstrated their advantageous properties by using single-molecule fluorescence microscopy to study immobilized proteins under destabilizing conditions. Biotinylated ribonuclease H (RNase H) was labeled with a fluorescence resonance energy transfer (FRET) pair of fluorescent dyes and attached to the surface by a biotin-streptavidin linkage. FRET analysis demonstrated completely reversible denaturation/renaturation behavior upon exposure of the surface-immobilized proteins to 6 M guanidinium chloride (GdmCl) followed by washing in buffer. A comparison with bovine serum albumin (BSA) coated surfaces and linear PEO brush surfaces yielded superior performance in terms of chemical stability, inertness and noninteracting nature of the star-polymer derived films.

Introduction

Chemically designed surface coatings that can prevent unspecific protein adsorption are essential for various biotechnological applications. Besides impeding biofouling, e.g., in membrane applications, nonadherent surface properties present a key condition for single molecule studies with immobilized proteins, for protein microarrays, and cell assays. To specifically bind proteins or to interact in a biomimetic way with living cells, surfaces have to be modified toward specific biological recognition,[1-4] whereas at the same time avoiding uncontrolled adsorption that could lead to denaturation of proteins or unwanted activation of biological processes. While protein immobilization itself is easy to accomplish, it is of utmost importance in studies of protein folding and function that the interaction between the protein under study and the surface environment is minimized. Otherwise, the results may not reflect intrinsic properties of the protein, but rather artifacts due to surface interactions. A number of surface preparations have been developed for the purpose of protein resistance, such as self-assembled monolayers on gold,[5-7] glass,[8] silicon,[9] titanium and titanium oxide,[10-13] polyelectrolyte multilayer films[1,14] and

† Department of Macromolecular Chemistry and Organic Chemistry III, University of Ulm.
‡ Department of Biophysics, University of Ulm.
§ Department of Physics, University of Illinois at Urbana-Champaign.
|| Deutsches Wollforschungsinstitut an der RWTH Aachen und Institut für Technische und Makromolekulare Chemie der RWTH Aachen.

(1) Chluba, J.; Voegel, J.-C.; Decher, G.; Erbacher, P.; Schaaf, P.; Ogier, J. *Biomacromolecules* 2001, 2, 800−805.
(2) Huang, K.; Lee, B. P.; Ingram, D. R.; Messersmith, P. B. *Biomacromolecules* 2002, 3, 397−406.
(3) Lahiri, J.; Ostuni, E.; Whitesides, G. M. *Langmuir* 1999, 15, 2055−2060.
(4) Kantlehner, M.; Schaffner, P.; Finsinger, D.; Meyer, J.; Jonczyk, A.; Diefenbach, B.; Nies, B.; Hölzemann, G.; Goodman, S. L.; Kessler, H. *ChemBioChem* 2000, 1, 107−114.

(5) Ostuni, E.; Chapman, R. G.; Holmlin, R. E.; Takayama, S.; Whitesides, G. M. *Langmuir* 2001, 17, 5605−5620.
(6) Harder, P.; Grunze, M.; Dahint, R.; Whitesides, G. M.; Laibinis, P. E. *J. Phys. Chem. B* 1998, 102, 426−436.
(7) Schwendel, D.; Dahint, R.; Herrwerth, S.; Schloerholz, M.; Eck, W.; Grunze, M. *Langmuir* 2001, 17, 5717−5720.
(8) Yang, Z.; Galloway, J. A.; Yu, H. *Langmuir* 1999, 15, 8405−8411.
(9) Petrash, S.; Cregger, T.; Zhao, B.; Pokidysheva, E.; Foster, M. D.; Brittain, W. J.; Sevastianov, V.; Majkrzak, C. F. *Langmuir* 2001, 17, 7645−7651.
(10) Tosatti, S.; Michel, R.; Textor, M.; Spencer, N. D. *Langmuir* 2002, 18, 3537−3548.
(11) Michel, R.; Lussi, J. W.; Csucs, G.; Reviakine, I.; Danuser, G.; Ketterer, B.; Hubbell, J. A.; Textor, M.; Spencer, N. D. *Langmuir* 2002, 18, 3281−3287.
(12) Chapman, R. G.; Ostuni, E.; Yan, L.; Whitesides, G. M. *Langmuir* 2000, 16, 6927−6936.
(13) Chapman, R. G.; Ostuni, E.; Liang, M. N.; Meluleni, G.; Kim, E.; Yan, L.; Pier, G.; Warren, H. S.; Whitesides, G. M. *Langmuir* 2001, 17, 1225−1233.
(14) Elbert, D. L.; Herbert, C. B.; Hubbell, J. A. *Langmuir* 1999, 15, 5355−5362.

hydrogels.[15] Poly(ethylene oxide) (PEO) brushes have been especially recognized as biocompatible and resistant to protein adsorption[5,6,8,16-18] due to the hydrophilic but uncharged nature of the polymer. Still, these methods suffer either from insufficient protein repellence, preparation methods that are tedious and difficult to reproduce, or low surface functionality. Here, we focus on a versatile, easily applicable functional surface coating for immobilization of proteins, prepared with a dense, ultrathin network formed from star shaped PEO molecules, linked at their chain ends via urea groups. We address the question as to which extent destabilization of the natural folded conformation can be avoided by the proper choice of surface linkage. Denaturation and renaturation of a protein, RNase H, specifically linked via biotin/streptavidin to ultrathin, cross-linked star polymer layers was studied by attaching a donor–acceptor pair of fluorescent dye molecules at specific locations along the polypeptide chain so that the two dyes are in close proximity in the folded structure and further apart in the unfolded chain. The strong distance dependence of fluorescence resonance energy transfer (FRET) enables direct insights into the folding dynamics of the polypeptide chain.[19-22] Our cross-linked star PEO surfaces showed superior performance compared with surfaces coated with linear PEO chains or pre-physisorbed proteins.

Single-Molecule Studies of Protein Folding

To be biologically active, the nascent polypeptide chain folds into a specific three-dimensional structure after biosynthesis. Due to the many possible internal degrees of freedom of the polymer, a large number of microscopic pathways exist that connect the vast number of unfolded conformations with the much smaller ensemble of native, folded conformations.[23] Therefore, protein folding is an inherently heterogeneous process. In recent years, the concept of a funnel-shaped conformational energy landscape has become prevalent, in which the folding polypeptide chain is guided toward the thermodynamic free energy minimum, encountering barriers and experiencing a gradual loss of enthalpy and entropy.[24] Single-molecule studies can provide direct experimental evidence on the folding/unfolding pathways in the complex energy landscape of proteins. In recent years, FRET analysis has been applied in two-color confocal fluorescence studies on proteins diffusing freely in solution.[21,22,25] In this method, the observation time is limited to the time it takes for a molecule to diffuse through the detection volume, which is on the order of milliseconds. To gain information about slower processes, for example in studies of folding intermediates, it is necessary to immobilize the proteins. Various immobilization strategies have been reported for single-molecule spectroscopy, including trapping

in porous polymer matrixes,[26,27] unspecific adsorption to surfaces[28,29] as well as specific adsorption via complex coordination of His-tagged proteins,[30,31] by using biotin/(strept-)avidin coupling[32,33] or charge interactions.[34,35] An immobilization technique, yielding minimal interaction with the environment, has been shown to be the trapping of proteins in surface-bound vesicles of ~100 nm diameter, in which they can freely diffuse within a limited volume.[36,37] In this approach, however, it is not simple to control the solution conditions inside the vesicles in situ. The star polymer layers introduced here offer an easy and more versatile alternative since they are prepared by spin coating a solution of isocyanate terminated, six-arm star polymers from aqueous THF onto amino-functionalized substrates. They were examined for unspecific adsorption of RNase H and were found to be essentially nonadsorbing. In contrast to adsorbed biotinylated BSA films on hydrophilic glass and PEO brushes made with a small fraction (1%) of biotinylated PEO chains, unfolding/refolding of RNase H was completely reversible on the star polymer derived surfaces, and offered high chemical stability.

Experimental Section

Synthesis and Labeling of RNase H. Plasmid pJAL135C containing the gene of single cysteine mutant of RNase H was a generous gift from Prof. S. Kanaya (Osaka University, Osaka, Japan). The protein was overproduced in *Escherichia coli* HB101 and purified as described.[38,39] The RNase H molecules were subsequently labeled with Alexa Fluor 546-NHS (Molecular Probes) and biotin-NHS (Sigma-Aldrich St. Louis, MO). For FRET measurements, a mutant of RNase H was constructed that had cysteine residues at positions 3 and 135, the thiol side chains of which were labeled with the dye FRET pair (Alexa 546/Alexa 647) by multimode coupling.[40]

Preparation of Isocyanate-Terminated Star PEO. Hydroxyl-terminated star polymers with 80% ethylene oxide and 20% propylene oxide as the backbone (number average molecular weight 12 000 g/mol; polydispersity index 1.15) were functionalized through reaction with a 12-fold molar excess of isophorone diisocyanate (IPDI) in a solvent-free process at 50 °C for 5 d.[41] The excess of IPDI was

(15) Lee, K. Y.; Mooney, D. J. *Chem. Rev.* **2001**, *101*, 1869−1877.
(16) Harris, J. M.; Zalipsky, S. *Poly(ethylene glycol): Chemistry and Biological applications*; American Chemical Society: Washington DC, 1997.
(17) Halperin, A. *Langmuir* **1999**, *15*, 2525−2533.
(18) Szleifer, I. *Biophys. J.* **1997**, *72*, 595−612.
(19) Kelley, A. M.; Michalet, X.; Weiss, S. *Science* **2001**, *292*, 1671−1672.
(20) Weiss, S. *Nat. Struct. Biol.* **2000**, *7*, 724−729.
(21) Deniz, A. A.; Laurence, T. A.; Beligere, G. S.; Dahan, M.; Martin, A. B.; Chemla, D. S.; Dawson, P. E.; Schultz, P. G.; Weiss, S. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 5179−5184.
(22) Schuler, B.; Lipman, E. A.; Eaton, W. A. *Nature* **2002**, *419*, 743−747.
(23) Frauenfelder, H.; Wolynes, P. G. *Phys. Today* **1994**, 58−64.
(24) Socci, N. D.; Onuchic, J. N.; Wolynes, P. G. *J. Chem. Phys.* **1996**, *104*, 5860−5868.
(25) Schwille, P.; Kettling, U. *Curr. Opin. Biotechnol.* **2001**, *12*, 382−386.
(26) Dickson, R. M.; Cubitt, A. B.; Tsien, R. Y.; Moerner, W. E. *Nature* **1997**, *388*, 3558−3358.
(27) Wiedenmann, J.; Schenk, A.; Rocker, C.; Girod, A.; Spindler, K. D.; Nienhaus, G. U. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 11 646−11 651.
(28) Ha, T.; Glass, J.; Enderle, T.; Chemla, D. S.; Weiss, S. *Phys. Rev. Lett.* **1998**, *80*, 2093−2096.
(29) Wazawa, T.; Ishii, Y.; Funatsu, T.; Yanagida, T. *Biophys. J.* **2000**, *78*, 1561−1569.
(30) Noji, H.; Yasuda, R.; Yoshida, M.; Kinosita, K. *Nature* **1997**, *386*, 299−302.
(31) Ha, T.; Ting, A. Y.; Liang, J.; Caldwell, W. B.; Deniz, A. A.; Chemla, D. S.; Schultz, P. G.; Weiss, S. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 893−898.
(32) Wennmalm, S.; Edman, L.; Rigler, R. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 10 641−10 646.
(33) Kim, H. D.; Nienhaus, G. U.; Ha, T.; Orr, J. W.; Williamson, J. R.; Chu, S. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 4284−4289.
(34) Talaga, D. S.; Lau, W. L.; Roder, H.; Tang, J.; Jia, Y.; DeGrado, W. F.; Hochstrasser, R. M. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 13 021−13 026.
(35) Jia, Y.; Talaga, D. S.; Lau, W. L.; Lu, H. S. M.; DeGrado, W. F.; Hochstrasser, R. M. *Chem. Phys.* **1999**, *247*, 69−83.
(36) Boukobza, E.; Sönnenfeld, A.; Haran, G. *J. Phys. Chem. B* **2001**, *105*, 12 165−12 170.
(37) Rhoades, E.; Gussakovsky, E.; Haran, G. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 3197−3202.
(38) Kanaya, S.; Crouch, R. J. *J. Biol. Chem.* **1983**, *258*, 1276−1281.
(39) Kanaya, S.; Kohara, A.; Miyagawa, M.; Matsuzaki, T.; Morikawa, K.; Ikehara, M. *J. Biol. Chem.* **1989**, *264*, 11 546−11 549.
(40) Heyes, C. D.; Kobitski, A. Y.; Amirgoulova, E. V.; Nienhaus, G. U., submitted.
(41) Goetz, H.; Beginn, U.; Bartelink, C. F.; Gruenbauer, H. J. M.; Moeller, M. *Macromol. Mater. Eng.* **2002**, *287*, 223−230.

removed by short path distillation. Size exclusion chromatography of the product (star PEO) proved that no dimer or trimer formation took place.

Surface Preparation and Specific Immobilization of RNase H. Cleaning, activation and aminofunctionalization of substrates was carried out under cleanroom conditions. Substrates were cleaned through sonication in acetone (Selectipur, Merck, Haar Germany), 18.2 MΩ Millipore water and 2-propanol (Selectipur, Merck) for one minute each. After activation by an oxygen plasma, the substrates were aminofunctionalized under an inert gas atmosphere for 2 h in a solution of 0.2 mL *N*-[3-(trimethoxysilyl)propyl] ethylenediamine (Sigma-Aldrich, 97%) in 50 mL dry toluene. Then the substrates were washed thoroughly and stored under dry toluene until further usage. For spin coating, the substrates were placed on the spin coater, covered by the star polymer solution and then accelerated within 5 s to 2500 rpm for 40 s. The resulting films were stored overnight at ambient atmosphere for cross-linking. For biotinylation, biocytin was dissolved in 9 mL of deionized water. This solution was mixed with the star polymer solution in 1 mL of THF. Films were then prepared as described above.

PEO surfaces were formed following cleaning and activation of glass substrates with a commercial aminosilane, Vectabond (Vector Laboratories, Burlingame, CA) according to the protocol recommended by the manufacturer. 100 mg/mL PEO solutions in 50 mM Na_2CO_3 buffer (pH 8.2) were prepared from mPEG-SPA, MW 5000 (Nektar Therapeutics, Huntsville, AL), or a mixture of Biotin-PEG-NHS, MW 3400 (Nektar Therapeutics) and mPEG-SPA, MW 5000 with 1% PEO-biotin by weight. PEO was reacted with the Vectabond amino-functionalized surface for 1 h in the dark. After completion of the reaction, samples were thoroughly washed with 18.2 MΩ Millipore water.

To form BSA covered surfaces, fluorescent contaminations were removed from untreated glass coverslips by brief exposure to an open flame. The surface was incubated with a 1 mg/mL BSA-biotin (Sigma-Aldrich) solution in 0.1 M sodium phosphate buffer (pH 7.4) for 10 min, then washed with the same buffer and used immediately afterward.

Biotinylated BSA/PEO/star polymer surfaces were exposed to 10 μg/mL streptavidin (Sigma-Aldrich) in 0.1 M sodium phosphate buffer (pH 7.4) for 10 min. Afterward, the surfaces were incubated with ~100 pM RNase H solution in buffer A (20 mM Tris-HCl, 100 mM KCl, 10 mM $MgCl_2$, pH 7.4) for 10 min. Finally, excess protein was washed with buffer A.

Single Molecule Measurements. Single-molecule microscopy was performed by using a homemade laser scanning confocal fluorescence microscope with Ar[+]/Kr[+] laser (modified model 164, Spectra Physics, Mountain View, CA) excitation. It is based on an inverted microscope (Axiovert 35, Zeiss, Göttingen, Germany) and has two separate detection channels for measurement of the emission in two spectral channels to enable FRET experiments. The setup is described in detail by Heyes et al.[40]

Results and Discussion

The cleaned and activated substrates were checked by scanning force microscopy (SFM, root-mean-square roughness ≤ 0.2 nm for 1 μm scans) and contact angle (below the detection limit with water). Ellipsometry measurements showed an increase in the SiO_x layer on the silicon of about 0.5 nm due to the activation step. The aminosilane layer exhibited a thickness between 1.1 and 1.5 nm and proved smooth when examined with SFM (root-mean-square roughness 0.6 nm for 1 μm scans).

Spin coating of the star polymers from aqueous THF solutions resulted in smooth, homogeneous films. The isocyanate terminated stars started to react with water already in solution. Some of the isocyanate groups converted into amines, which reacted with isocyanate groups to yield di-, tri-, and higher oligomers of stars (Figure 1). Consequently, during spin coating, oligomers
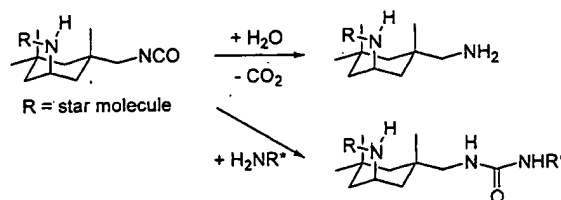
Figure 1. Scheme of the cross-linking reaction of the polymer.
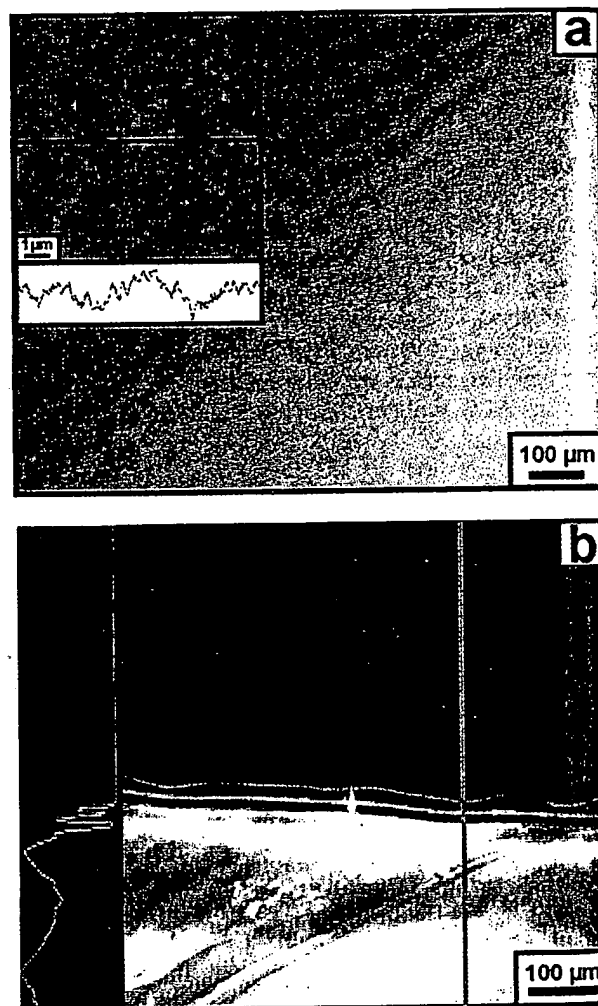
Figure 2. Images of star polymer derived surfaces. (a) optical microscopy (DIC) and scanning force microscopy (height image; profile height is 2 nm) highlight the smoothness of the polymer films. (b) fluorescence microscopy image of a star polymer covered substrate half-dipped into a polystyrene solution. Labeled streptavidin adsorbs unspecifically on polystyrene but not on the star polymer. Fluorescence intensities are ~1000 s[−1] for the star polymer covered area (equal to background intensity) and >15 000 s[−1] for the polystyrene covered area.

coexisted with monomers containing amine groups and unmodified monomers. Unreacted isocyanate groups can bind covalently to the amine groups on the aminosilanized substrate surface. Because of partial hydrolysis and cross-linking in solution, the deposited layer then reacts quickly to a highly cross-linked network. With the star polymer concentration chosen for these experiments (1 mg/mL), a film thickness of 5 ± 0.5 nm was measured by ellipsometry on silicon substrates corresponding to at most three monolayers. The contact angle with water was determined by sessile drop measurements (advancing contact angle) as 52° ± 3° (Θ_{adv}) and by captive bubble measurements after storing the samples for 12 h in deionized water (receding
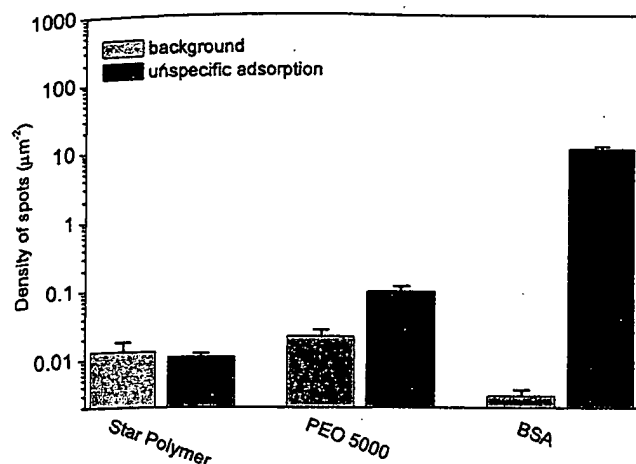
Figure 3. Analysis of the density of fluorescent spots on freshly prepared surfaces and after exposure to RNase H molecules labeled with a single fluorophore. Significantly higher levels of unspecific adsorption were observed on BSA than on PEO surfaces. The star polymer surfaces showed negligible unspecific adsorption.

contact angle) as $45° \pm 3°$ (Θ_{rec}). Figure 2a shows the exquisite smoothness of the film, as examined by optical microscopy and SFM.

To examine the resistance of the star polymer films to unspecific protein adsorption, the star polymer covered substrates were dipped halfway into a polystyrene solution in toluene. Polystyrene surfaces are known for strong unspecific adsorption of proteins.[38] The so-created half polystyrene covered samples were immersed into protein solutions (streptavidin and avidin, labeled with fluorescent dyes) in different buffer systems (pH 5, pH 7.4, and pH 9.5). The proteins adsorbed onto the polystyrene, but not at all onto the star polymer coating (Figure 2b). In a control experiment, unspecific protein adsorption on plain, aminosilanized wafers was shown to be high (not shown). Therefore, prevention of unspecific protein adsorption on the star polymer surfaces was demonstrated from pH 5 to 9.5—the pH range important for functional biomolecules.

Unspecific adsorption was studied at the single-molecule level with unbiotinylated star-polymer surfaces and, for comparison, with physisorbed biotinylated BSA surfaces and with unbiotinylated PEO brush surfaces. All three samples were simultaneously incubated with the same ~15 nM solution of single-dye labeled RNase H in buffer A for 10 min and then thoroughly washed with buffer A. Because streptavidin was absent from the solution, specific binding to the biotinylated BSA was excluded. Control experiments of the surface cleanliness were performed to ensure that the observed fluorescent spots were from labeled RNase H and not from contamination. The amount of unspecific adsorption was calculated from the density of spots (single molecules) on each of the surfaces. Typically, twenty images (18 × 18 μm) of each sample were examined to obtain statistical significance. The data in Figure 3 show that the BSA surfaces were the cleanest in terms of background fluorescence, but they also showed the highest unspecific adsorption. The star polymer derived surfaces and PEO surfaces were slightly more contaminated, but showed markedly lower amounts of unspecifically adsorbed RNase H proteins. In fact, the density of spots did not increase at all from the background density on the star polymer layers, underscoring their highly adsorption-resistant nature.
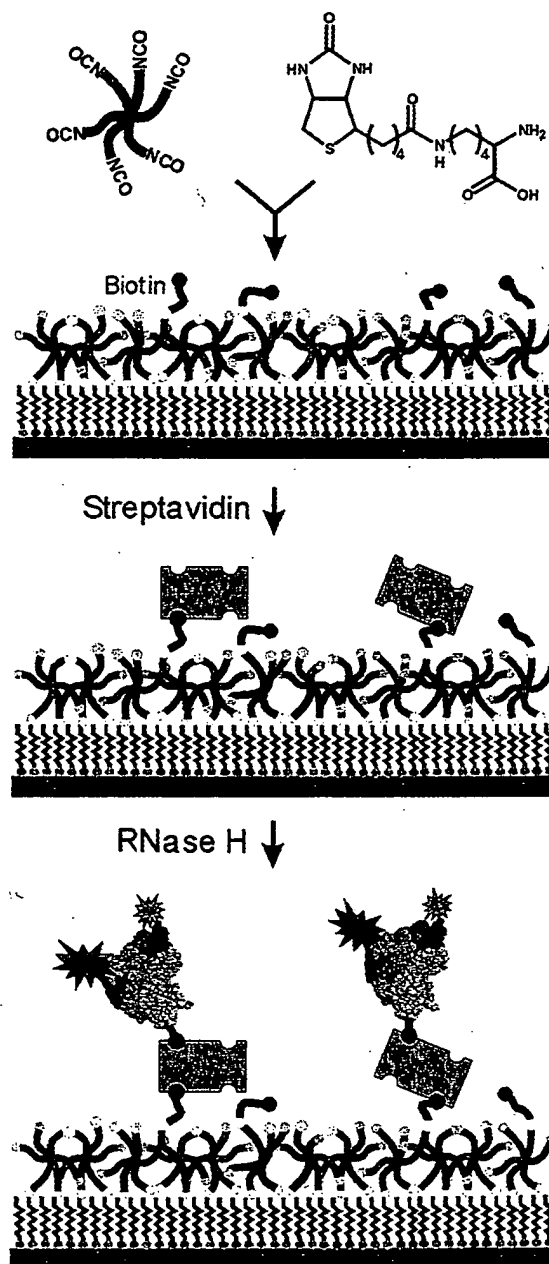
Figure 4. Preparation of biotinylated, star pre-polymer derived layers with subsequent specific binding of streptavidin and RNase H.

The star polymers were further modified with biotin anchors for specific attachment of RNase H molecules to the surface (Figure 4). To this end, biocytin was reacted with the isocyanate-terminated stars just prior to spin coating, yielding surfaces that were statistically decorated with biotin groups. These coatings were subsequently exposed to streptavidin (10 μg/mL in phosphate buffer for 10 min). Streptavidin is tetravalent to biotin, and biotinylated RNase H in buffer A was attached to the surface via the vacant binding sites on the already bound streptavidin.

Denaturation experiments were performed with RNase H molecules immobilized on the star polymer surface. To monitor the conformation of the polypeptide chain, a FRET pair of dye molecules was attached at positions 3 and 135 of the RNase H sequence. These locations were chosen such that of the dye molecules are in close proximity in the native conformation and significantly further apart in the denatured state. The FRET efficiency of each individual molecule can be calculated from
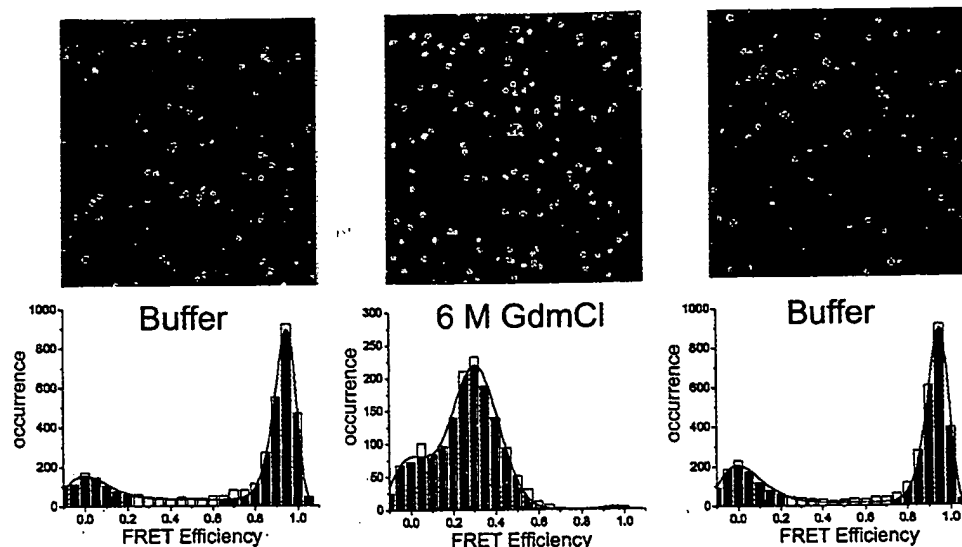
Figure 5. Two-color (red/green) images (top) and histograms of the number of molecules as a function of their FRET efficiencies (bottom) of RNase H bound to star polymer derived surfaces. One complete denaturation/renaturation cycle is shown, starting from the initial preparation in buffer solution (left) via the denatured state in 6 M GdmCl (middle) back to buffer solution (right). The changes in the distributions of FRET efficiencies indicate that the protein molecules unfold (bright green, $E \approx 0.3$) and refold (red, $E \approx 1$) completely reversibly. The peak at zero FRET efficiency (dark green) is due to RNase H molecules without a, or with an already bleached, red dye molecule.
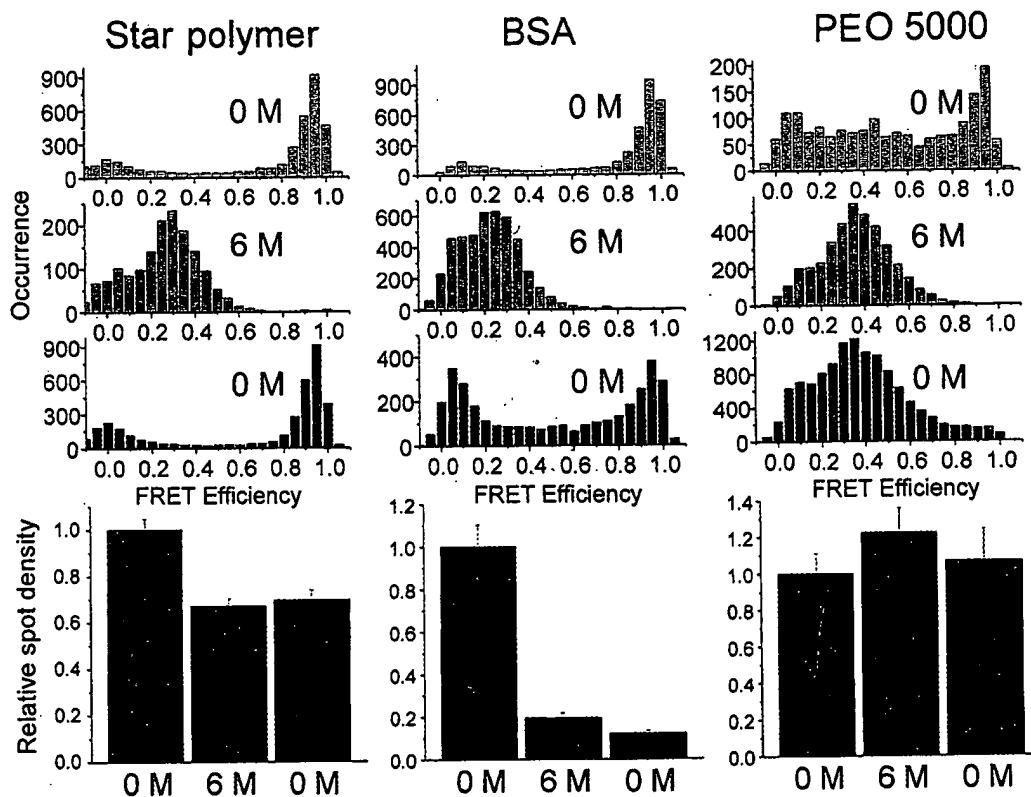
Figure 6. FRET efficiency histograms (top) and relative protein densities on the surface (bottom) during one denaturation/renaturation cycle (0 M, 6 M and finally 0 M GdmCl) of RNase H bound to a star polymer derived surface (left column), adsorbed BSA surface (middle column) and PEO 5000 polymer brush surface (right column). On PEO 5000 brush surfaces, protein denaturation is essentially irreversible. On BSA, labeled RNase H disappears from the surface under denaturing conditions, presumably together with the BSA layer. In contrast, RNase H can be unfolded and refolded completely reversibly on star polymer derived surfaces. Moreover, the spot density on the star polymer surface does not change after treatment with denaturant.

the intensities of photons in the red (acceptor) and green (donor) channels, I_A and I_D, respectively

$$E = \frac{I_A}{I_A + \gamma I_D}$$

The correction factor γ accounts for the difference in detection

efficiency between the two channels. The FRET efficiency varies with the inverse sixth power of the dye-to-dye separation and is thus exquisitely sensitive to distance variations due to structural changes of the protein. Denaturation and subsequent renaturation of surface-bound RNase H was performed by variation of the concentration of guanidinium chloride (GdmCl) as a denaturing agent. Figure 5 (top) shows scan images taken

under different solvent conditions with excitation at 514 nm. Each spot represents an individual RNase H molecule. In the initial scan, performed in buffer, the majority of spots emitted photons predominantly into the red spectral channel (sensitive only to emission from the red dye), reflecting a large FRET efficiency ($E \approx 1$), and thus a close proximity of the dyes in the folded molecules. Upon exchange of the buffer with buffer containing 6 M GdmCl, most photons emitted from the spots were detected in the green spectral channel (sensitive to emission from the green dye) because of the lower FRET efficiency ($E \approx 0.3$) arising from the larger (average) dye-to-dye separation in the unfolded state. After re-exchange of the denaturant with buffer A, the spots turned red again, implying that the proteins refolded into the compact, high-FRET conformation. A small population of green spots prior to denaturation and a slightly increased number of green spots after renaturation represented molecules without red acceptor dye, because either it was lacking in the first place or it was photobleached during the experiment. These qualitative results were confirmed by a thorough quantitative analysis of the FRET efficiencies in Figure 5 (bottom), which shows complete refolding on the star polymer surfaces.

Figure 6 shows the comparison of the experiment presented in Figure 5 for the three surfaces examined. Together with the FRET data, we have also plotted the measured spot densities, normalized to unity for the first scan (in buffer). For the star polymer surfaces and the PEO brushes, the spot densities are relatively unaffected by the harsh chemical treatment with 6 M GdmCl. PEO brushes, however, are seen to completely prevent renaturation of the proteins after unfolding. Even in the first scan (in buffer), many molecules show low FRET efficiencies on linear PEO brush surfaces, due to partial or complete unfolding of the protein molecules. Most likely, the extended polypeptide chains interact and intermingle with the long PEO chains so that they cannot refold into their native conformation. Alternatively, bound RNase H may penetrate the flexible PEO brushes and interact with the underlying aminosilane. In any case, this process is apparently prevented by the extensively cross-linked star polymers. On the BSA surfaces, a substantial fraction of RNase H refolds, as seen from the recurrence of the maximum at high FRET efficiencies. Yet, the broad pedestal at intermediate FRET efficiencies suggests that some fraction of the molecules did not refold properly. Moreover, measurement of the spot (protein) density on the surfaces revealed that the protein concentration on BSA surfaces was reduced significantly under 6 M GdmCl. This observation implies that a large fraction of RNase H proteins was removed by treatment with GdmCl, most likely together with the BSA layer physisorbed to the glass.

Conclusions

Ultrathin, smooth layers from isocyanate terminated star polymers on glass substrates were shown to be extremely resistant to unspecific adsorption of proteins while at the same time suitable for easy chemical modification. Application by spin coating offers a simple procedure for the preparation of minimally interacting surfaces that are functionalized by suitable linker groups to immobilize proteins in their active natural conformation. These coatings form a versatile platform for biofunctional and biomimetic surfaces and single-molecule fluorescence microscopy studies on immobilized proteins. In single-molecule denaturation/renaturation experiments with RNase H molecules specifically attached to the star polymers, complete reversibility of this process was observed, implying minimal interaction between the protein and the surface. A comparison with adsorbed BSA and PEO brush surfaces clearly demonstrated the superior quality of the star polymer layers.

Journal of the American Chemical Society

Exhibit 22

A method for global protein expression and antibody screening on high-density filters of an arrayed cDNA library

Konrad Büssow*, Dolores Cahill, Wilfried Nietfeld, David Bancroft, Eberhard Scherzinger, Hans Lehrach and Gerald Walter

Max Planck Institute for Molecular Genetics, Ihnestr. 73, D-14195 Berlin, Germany

DDBJ/EMBL/GenBank accession no. AF074376

ABSTRACT

We have developed a technique to establish catalogues of protein products of arrayed cDNA clones identified by DNA hybridisation or sequencing. A human fetal brain cDNA library was directionally cloned in a bacterial vector that allows IPTG-inducible expression of His$_6$-tagged fusion proteins. Using robot technology, the library was arrayed in microtitre plates and gridded onto high-density *in situ* filters. A monoclonal antibody recognising the N-terminal RGSH$_6$ sequence of expressed proteins (RGS·His antibody, Qiagen) detected 20% of the library as putative expression clones. Two example genes, GAPDH and HSP90α, were identified on high-density filters using DNA probes and antibodies against their proteins.

For construction of the human expression library hEx1, cDNA was prepared from fetal brain poly(A)$^+$ RNA by oligo(dT)-priming (Superscript Plasmid System, Life Technologies). Products were size-fractionated by gel filtration and directionally (*SalI–NotI*) cloned into a modified pQE-30 (Qiagen) vector for IPTG-inducible expression of His$_6$-tagged fusion proteins (pQE-30NST, GenBank accession no. AF074376). *Escherichia coli* SCS1 cells (Stratagene) carrying the plasmid pSE111 with the *lacI^Q* repressor and the *argU* gene for a rare arginine tRNA (1) were transformed by electroporation. PCR analysis of 96 clones revealed an average insert size of ~1.5 kb (range 0.5–5.0 kb).

The library was plated onto 2×YT-AKG agar plates (230 mm × 230 mm Nunc Bio Assay Dishes containing 2×YT agar, 100 µg/ml ampicillin, 15 µg/ml kanamycin and 2% glucose) and grown at 37°C overnight. Using a picking/gridding robot (2), 193 536 colonies were picked into 384-well microtitre plates (Genetix) containing 2×YT-AKG medium supplemented with freezing mix (0.4 mM MgSO$_4$, 1.5 mM Na$_3$-citrate, 6.8 mM (NH$_4$)$_2$SO$_4$, 3.6% glycerol, 13 mM KH$_2$PO$_4$, 27 mM K$_2$HPO$_4$, pH 7.0). Bacteria were grown in microtitre wells at 37°C overnight, and 9216 or 27 648 clones were gridded onto 222 mm × 222 mm filter membranes in a duplicate pattern (Fig. 1). Nylon filters (Hybond-N$^+$, Amersham) were gridded for DNA hybridisations and processed as described (3). For protein analysis, polyvinylidene difluoride (PVDF)

filters (Hybond-P, Amersham) were gridded, incubated on 2×YT-AKG agar plates at 30°C overnight and induced for protein expression for 3 h at 37°C on agar plates containing 1 mM IPTG. These protein filters were processed on pre-soaked blotting paper, i.e., denatured in 0.5 M NaOH, 1.5 M NaCl for 10 min, neutralised for 2× 5 min in 1 M Tris–HCl, pH 7.5, 1.5 M NaCl and incubated for 15 min in 2× SSC. Filters were air-dried and stored at room temperature.

For global protein expression, high-density filters were screened with the monoclonal antibody RGS·His (Qiagen). This antibody recognises the N-terminal sequence RGSH$_6$ of fusion proteins over-expressed from pQE-30 vectors and labelled ~20% of the hEx1 clones (Fig. 1A). Negative clones have inserts in incorrect reading frames with stop codons leading to short polypeptides that cannot fold into stable structures and are degraded within the host cell (4). Two example proteins, GAPDH (35.9 kDa, Swiss-Prot P04406) and HSP90α (84.5 kDa, Swiss-Prot P07900) were chosen for detailed analysis. A set of three DNA filters (80 640 clones) were screened with cDNA probes. Two hundred and six (0.26%) clones were positive with a human GAPDH probe (Fig. 1B), and 56 (0.07%) clones were identified with a human HSP90α probe. About 25% of these clones were positive with the RGS·His antibody. To confirm the expression of GAPDH or HSP90α proteins by these clones, protein filters were screened with antibodies against GAPDH (Fig. 1C) or HSP90α, respectively. Fifty-seven percent of the GAPDH and 72% of the HSP90α clones detected by the RGS·His antibody were also positive with the protein-specific antibodies. Sequence analysis showed that the remaining clones had inserts in an incorrect reading frame or expressed truncated GAPDH which reacted poorly with the GAPDH antibody.

In turn, 100% of the anti-GAPDH- but only 35% of the anti-HSP90α-positive clones were detected by the RGS·His antibody. All RGS·His-negative HSP90α clones had inserts in incorrect reading frames but nevertheless expressed proteins detected by the HSP90α antibody on western blots (data not shown). This indicates HSP90α molecules without a His$_6$ tag, suggesting translational start sites within cDNA inserts. Three anti-HSP90α positive clones contained inserts that were not recognised by the cDNA probe and turned out to be unrelated

*To whom correspondence should be addressed. Tel: +49 30 8413 1614; Fax: +49 30 8413 1380; Email: buessow@mpimg-berlin-dahlem.mpg.de
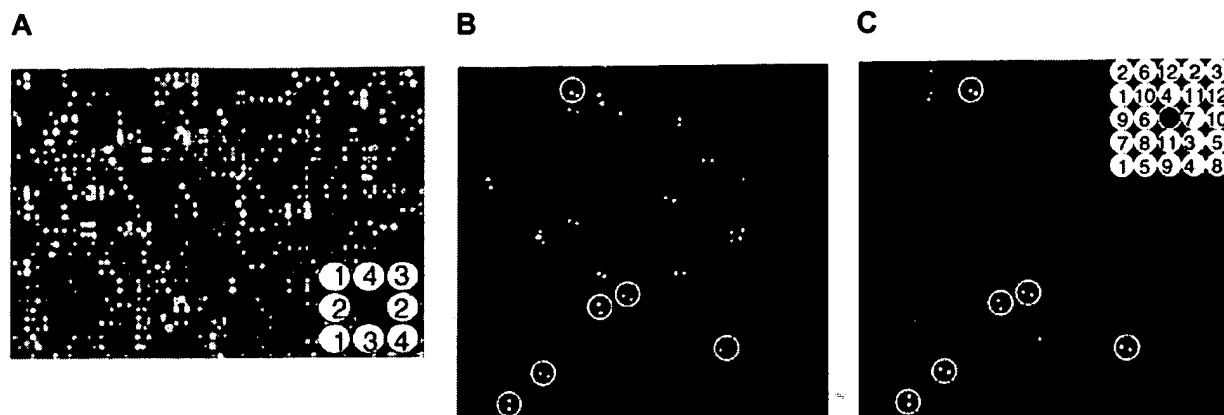
A **B** **C**

Figure 1. Identification of cDNA clones expressing recombinant fusion proteins on high-density filters. (A) RGS·His antibody detection (gridding pattern of 3×3 surrounding ink guide dots as shown in lower right corner). (B) DNA hybridisation with a GAPDH cDNA probe, as described (3). (C) Screening with a polyclonal anti-GAPDH antibody (corresponding sections of filters featuring 5×5 gridding patterns as shown in upper right corner; identical clones are circled). Before antibody screening, filters were soaked in ethanol, bacterial debris was wiped off in TBST-T (20 mM Tris–HCl pH 7.5, 0.5 M NaCl, 0.1% Tween 20, 0.5% Triton X-100), followed by washing 2× 10 min in TBST-T and 10 min in TBS. Filters were blocked in blocking buffer (3% non-fat, dry milk powder in TBS, 150 mM NaCl, 10 mM Tris–HCl, pH 7.5) for 1 h and incubated with antibody for 2 h (1:2000 diluted monoclonal RGS·His, Qiagen, or 50 ng/ml monoclonal anti-HSP90α, Transduction Laboratories, Lexington) or for 16 h (1:5000 diluted rabbit anti-GAPDH). After washing for 2× 10 min in TBST-T and 10 min in TBS, filters were incubated with alkaline phosphatase-conjugated anti-mouse or anti-rabbit Ig (Pierce) for 1 h, washed 3× 10 min in TBST-T, 10 min in TBS and 10 min in AP buffer (1 mM MgCl₂, 0.1 M Tris–HCl, pH 9.5) and incubated in 0.25 mM Attophos (JBL Scientific, San Luis Obispo) in AP buffer for 5 min. Filters were illuminated with long-wave UV light and images were taken using a high resolution CCD detection system. Image analysis was done using Xdigitise software (written by Huw Griffith) which is available on request.

sequences. These sequences were analysed using BESTFIT (Wisconsin Package Version 9.1, Genetics Computer Group, Madison) but no common motifs of significant homology were found. This limited antibody specificity is not surprising as it reflects cross-reactivity which is not usually tested against a whole library of proteins as in our method.

The main advancement of our technique over existing technology (e.g. λgt11 libraries; 5) is its high-throughput link between DNA sequence information and protein expression as a resource for unlimited future use. Having screened a library for protein expression once, we can always go back and identify products of new genes as they are discovered, attributing first functional information to them. Based on screenings with the RGS·His antibody, 37 830 putative expression clones were re-arrayed into new microtitre plates, and high-density protein and DNA filters were prepared and are available from the Resource Centre of the German Human Genome Project (http://www.rzpd.de).

The main technical problems of our approach are inherent in cDNA library and filter hybridisation technology. Oligo(dT)-primed cDNA is biased towards 3′-ends of genes, and, subject to insert size, N-terminal parts of larger proteins are often missing. To include a maximum number of epitopes for antibody screening, complementary random-primed cDNA libraries should be used. Quantification of signal intensities on filters is largely based on arbitrary thresholds for manual or automated image analysis. Therefore, our approach is exclusively based on positive clones to be confirmed by sequencing and/or protein characterisation.

We envisage two main fields of application for our method. First, catalogues of protein products can be established for different tissues and developmental stages. As these proteins are expressed from arrayed cDNA clones, their identity can easily be

checked by high-throughput gene identification techniques (e.g. oligonucleotide fingerprinting; 6). Therefore, gene expression patterns of normal and diseased tissues can be translated to the protein level, keeping a direct link to already existing DNA sequence data. Second, our method should also enable high-throughput analysis of antibody specificity and other protein–protein interaction or ligand-receptor systems (7), including non-protein molecules.

ACKNOWLEDGEMENTS

REFERENCES

1 Brinkmann,U., Mattes,R.E. and Buckel,P. (1989) *Gene*, 85, 109–114.
2 Lehrach,H., Bancroft,D. and Maier,E. (1997) *Interdisciplinary Science Reviews*, 22, 37–44.
3 Maier,E., Crollius,H.R. and Lehrach,H. (1994) *Nucleic Acids Res.*, 22, 3423–3424.
4 Anon. (1997), *The QIAexpressionist*. 3rd Ed. Qiagen, Hilden, Germany, pp. 31–35.
5 Young,R.A. and Davis,R.W. (1983) *Proc. Natl Acad. Sci. USA*, 80, 1194–1198.
6 Meier,E.S., Lange,J., Gerst,H., Herwig,R., Schmitt,A., Freund,J., Elge,T., Mott,R., Herrmann,B. and Lehrach,H. (1998) *Nucleic Acids Res.*, 26, 2216–2223.
7 Sparks,A.B., Hoffman,N.G., McConnell,S.J., Fowlkes,D.M. and Kay,B.K. (1996) *Nature Biotechnol.*, 14, 741–744.

Exhibit 23

A Human cDNA Library for High-Throughput Protein Expression Screening

Konrad Büssow,[1] Eckhard Nordhoff, Christine Lübbert, Hans Lehrach, and Gerald Walter

Max Planck Institute of Molecular Genetics, Ihnestrasse 73, 14195 Berlin, Germany

Received October 27, 1999; accepted January 21, 2000

We have constructed a human fetal brain cDNA library in an *Escherichia coli* expression vector for high-throughput screening of recombinant human proteins. Using robot technology, the library was arrayed in microtiter plates and gridded onto high-density filter membranes. Putative expression clones were detected on the filters using an antibody against the N-terminal sequence RGS-His$_4$ of fusion proteins. Positive clones were rearrayed into a new sublibrary, and 96 randomly chosen clones were analyzed. Expression products were analyzed by SDS–PAGE, affinity purification, matrix-assisted laser desorption/ionization-time-of-flight mass spectrometry, and the determined protein masses were compared to masses predicted from DNA sequencing data. It was found that 66% of these clones contained inserts in a correct reading frame. Sixty-four percent of the correct reading frame clones comprised the complete coding sequence of a human protein. High-throughput microtiter plate methods were developed for protein expression, extraction, purification, and mass spectrometric analyses. An enzyme assay for glyceraldehyde-3-phosphate dehydrogenase activity in native extracts was adapted to the microtiter plate format. Our data indicate that high-throughput screening of an arrayed protein expression library is an economical way of generating large numbers of clones producing recombinant human proteins for structural and functional analyses. © 2000 Academic Press

INTRODUCTION

Cellular functions are controlled by the networked expression of gene catalogues. Functional network analysis requires the parallel expression and characterization of large numbers of gene products. Structural analysis provides clues to biochemical functions of unknown proteins (Hwang *et al.*, 1999; Zarembinski *et al.*, 1998). Genome analysis by DNA hybridization and sequencing has become a highly automated process (Lehrach *et al.*, 1997). In contrast, the individual-

ity of protein molecules demands highly customized procedures for their expression. Automation of these procedures requires systems that allow the efficient handling of large numbers of clones representing many different proteins. Bacterial systems are easy to manage but the expression of eukaryotic proteins can be problematic, due to aggregation, formation of insoluble inclusion bodies, and/or degradation of the expression product (Hockney, 1994; Makrides, 1996). Eukaryotic systems suffer from lower yields of heterologous protein (e.g., *Saccharomyes cerevisiae;* Buckholz and Gleeson, 1991), high demands on sterility (e.g., mammalian systems; Aruffo, 1997; Kingston *et al.*, 1997), or time-consuming cloning procedures (e.g., Baculovirus system; Miller, 1993).

We have shown that automated technology can be used for high-throughput protein expression screening (Büssow *et al.*, 1998; Lueking *et al.*, 1999). Mammalian cDNA libraries are directly cloned into bacterial expression vectors, circumventing the subcloning of individual protein-coding sequences. In a first screening round, putative protein-expressing clones are identified on high-density filters using antibodies against a vector-encoded tag sequence. The detected clones are rearrayed into a smaller sublibrary. In a second round, small-scale protein expression is performed in microtiter plates. Products are analyzed for size, yield, homogeneity, and solubility using SDS–PAGE, affinity purification, and matrix-assisted laser desorption/ionization–time-of-flight mass spectrometry (MALDI–TOF–MS). Expression levels of large numbers of clones are assessed in parallel to find the most suitable for high-throughput structural analyses by X-ray crystallography or NMR and functional screening. In a third round, protein function is also assayed in the microtiter plate format. As an example, bacterial lysates of 96 clones were screened for expression of glyceraldehyde-3-phosphate dehydrogenase (GAPDH) activity. In summary, our multistep screening approach enables the generation of an expression clone catalogue of human proteins as a resource for structural and functional genomic analyses.

[1] To whom correspondence should be addressed. Telephone: +49-30-8413-1614. Fax: +49-30-8413-1128. E-mail: buessow@molgen.mpg.de.

MATERIALS AND METHODS

cDNA library and protein expression screening on high-density filters. A cDNA library (hEx1) from human fetal brain tissues was cloned in the expression vector pQE30NST (GenBank Accession No. AF074376). High-density protein filters were prepared and were screened with the RGS · His antibody (Qiagen), as described (Büssow *et al.*, 1998). In total, 193,536 clones of the hEx1 library were picked and labeled according to the RZPD nomenclature (http://www.rzpd.de). Clone names contain the library number MPMGp800 as a prefix.

PCR amplification and sequencing. cDNA inserts were amplified using primers pQE65 (TGAGCGGATA ACAATTTCAC ACAG) and pQE276 (GGCAACCGAG CGTTCTGAAC) at an annealing temperature of 65°C. PCR products were tag-sequenced using primer pQE65.

Protein expression and nickel chelate affinity chromatography in microtiter plates. Proteins were expressed in 1-ml cultures in deep-well microtiter plates, and protein extracts were obtained as described (Lueking *et al.*, 1999). Twenty-five microliters of 50% Ni-NTA agarose was added to protein extracts obtained under denaturing conditions, and His$_6$-tagged proteins were bound by shaking for 1 h in a microtiter plate shaker. The agarose beads were washed three times by resuspending them in Buffer C (8 M urea, 0.1 M NaH$_2$PO$_4$, 0.01 M Tris, pH 6.3), shaking for 5 min, and removal of liquid on the vacuum filtration manifold. Buffer C was removed by washing four times with 200 µl of 5 mM Tris-HCl, pH 8.0. Proteins were eluted by adding 100 µl of 35% acetonitrile, 0.1% TFA, shaking for 10 min, followed by centrifugation at 2000 rpm for 2 min, and collection of eluates in a fresh 96-well microtiter plate. Five microliters of the eluates was analyzed by SDS-PAGE, and 0.5-µl aliquots were subjected to MALDI-TOF-MS analysis.

MALDI-TOF-MS analyses. Aliquots (0.5 µl) of protein eluates were loaded onto a Bruker Scout-384 MALDI sample support (384 sample positions arranged according to the microtiter plate format), followed by addition of 0.5 µl sinapic acid matrix solution (saturated in 35% acetonitrile). The samples were deposited onto the central positions E7–E18. . .L7–L18. In addition, a protein calibration standard containing 0.5 pmol horse heart cytochrome c and 1 pmol human carbonic anhydrase was placed between the sample positions H12, I12, H13, and I13. All samples were analyzed on a Bruker Scout 384 Biflex III MALDI-TOF mass spectrometer in linear operational mode using externally determined calibration constants. Exclusively positively charged ions were detected, and 100–150 single-shot spectra were accumulated for improved signal-to-noise ratio. Before the analysis, the instrumental parameters were optimized for good signal resolution in the mass range 10–30 kDa using the protein calibration standard, and external calibration constants were determined using the molecular ion signals of cytochrome c and human carbonic anhydrase I. If indicated by the above measurements, proteins contained in 10 µl eluate were neutralized and reduced by addition of 2 µl containing 500 mM Tris-HCl, pH 7.5, 50 mM DTT and incubated at 55°C for 30 min. Aliquots (0.5 µl) of these solutions were deposited onto the MALDI sample support followed by 0.5 µl sinapic acid matrix solution containing also 2.5% TFA. After solvent evaporation, these samples were analyzed as described above.

GAPDH assay. The GAPDH assay described by Heinz and Freimüller (1982) was adapted to the microtiter format and performed in duplicate. One hundred fifty microliters of assay mix (33 mM TEA-HCl, 0.23 mM NADH, 6.7 mM MgSO$_4$, 1 mM ATP, 3 mM glycerate 3-phosphate, 3.8 mM L-cysteine) was added to 1-µl soluble protein fractions diluted 1:10, in 96-well microtiter plates (Microtest III, Falcon). The decrease of A_{340} was measured with a microtiter plate photometer (Spectramax 250, Molecular Devices). One unit of GAPDH catalyzes the reduction of 1 µmol of 1,3-diphosphateglycerate to D-glyceraldehyde-3-phosphate per minute.

RESULTS

A human fetal brain cDNA expression library (hEx1) was constructed in the vector pQE30NST, which allows the expression of fusion proteins with the N-terminal sequence RGS-His$_6$ (Büssow *et al.*, 1998). Briefly, 193,536 clones were picked into 384-well microtiter plates (plates 1 to 504 of hEx1) and gridded as high-density protein filters using a robotic system (Lehrach *et al.*, 1997). These clone arrays were screened for putative protein expression clones using the monoclonal antibody RGS · His (Qiagen), which recognizes the N-terminal sequence RGS-His$_6$ of recombinant expression products. The antibody preferentially labels clones containing a cDNA insert in-frame with RGS-His$_6$. In alternative reading frames, stop codons cause the expression of short and unstable products that are degraded in the *Escherichia coli* host cell (Gottesman, 1996). A total of 37,830 (19.6%) clones were recognized by the RGS · His antibody, 67% of which were labeled with medium or high intensity. All positive clones were combined in a new library by rearraying in 99 × 384-well microtiter plates (labeled plates 505 to 603 of hEx1) using the same robotic system equipped with dedicated rearraying software (Büssow *et al.*, 1998).

cDNA inserts of 96 randomly chosen clones of the medium and high RGS · His signal intensity groups were sequenced. All 96 clones originated from plate 582 of the rearrayed hEx1 library. cDNA inserts were amplified by PCR and analyzed by 5'-tag sequencing. An average insert size of 1.5 kb was determined. 5'-tag sequences of 93 cDNA inserts were obtained and used to search SP-TrEMBL, the combined SWISS-PROT and TrEMBL protein database (Bairoch and Apweiler, 1998) using the program BLASTX (Altschul *et al.*, 1990). Fifty-nine sequences were found to match human proteins in this database (Table 1). Thirty-eight (64%) of those sequences matched the beginning of a human protein, suggesting that the complete coding region had been cloned (full-length clones). Thirty-nine (66%) of the 59 sequences were fused to the N-terminal sequence RGS-His$_6$ in the correct reading frame (RF+). Protein molecular masses were predicted for these clones by completing their 5'-tag sequences using the matching sequences in the database (Table 1), considering that the formyl group of the N-terminal formyl-methionine is removed in *E. coli* and that the resulting N-terminal methionine is usually not removed if it is followed by arginine (Sherman *et al.*, 1985).

Expression products of the same 96 clones were analyzed by SDS-PAGE of cellular protein extracts and by nickel chelate affinity purification, followed by both SDS-PAGE and MALDI-TOF-MS. Protein expression and all subsequent steps were performed in 96-well microtiter plates. Seventy-two (75%) of the total 96 clones expressed recombinant proteins detectable in SDS-PAGE. Thirty-five of the 39 in-frame clones produced RGS-His$_6$ tag fusion proteins of expected sizes,

TABLE 1

Protein Expression Properties of hEx1 Clones with Sequence Database Matches

Clone MPMGp 800...	SP-TrEMBL database match		First matched amino acid in database sequence	Reading frame (RF)	Predicted protein size (kDa)	Expressed protein size, measured by MALDI-TOF-MS (kDa)	Expressed protein size, estimated by SDS-PACE (kDa)*
	Accession No.	Protein name					
A06582	O00217	NADH-ubiquinone oxidoreductase 23 kDa subunit precursor (EC 1653)	1	–	—	8,602	12
A08582	P04687	Tubulin α-1 chain	174	+	33,859	33,865	35
A12582	P04765	Eukaryotic initiation factor 4A-I (EIF-4A-I)	12	+	47,808	47,819	50
A14582	Q11203	CMP-N-acetylneuraminate-B-1,4-galactoside α-2,3-sialyltransferase	234	+	18,825	18,831ᵇ	20
A16582	P13639	Elongation factor 2 (EF-2)	1	+	98,098	(18,458)	(97)
A18582	P49006	Marcks-related protein (MAC-MARCKS)	1	–	—	12,282	14
A20582	P56182	NNP-1 protein (D21S2056E)	240	+	27,850	27,857	34
A24582	P49006	Marcks-related protein (MAC-MARCKS)	1	–	—	10,052	13
C06582	Q15853	Upstream stimulatory factor 2	128	–	—	13,908	18
C10582	P36578	60S Ribosomal protein L1 (L4)	1	–	—	10,304	10
C12582	P49006	Marcks-related protein (MAC-MARCKS)	1	–	—	13,557	16
E02582	P43308	Translocon-associated protein, β subunit precursor (TRAP-β)	1	–	—	10,753	10
E04582	P14793	60S Ribosomal protein L40 (CEP52)	36	–	—	7,750	10
E10582	O75312	Zinc-finger protein ZPR1	1	+	54,299	54,334	55
E12582	P25111	40S Ribosomal protein S25	1	–	—	7,504	10
E14582	P04687	Tubulin α-1 chain	304	–	—	11,301	10
E18582	Q15560	Transcription elongation factor S-II	1	+	39,238	8,445	10
E20582	Q13885	β Tubulin	275	+	22,579	22,594	23
C02582	P14923	Junction plakoglobin	287	+	53,034	53,066	50
C04582	Q16478	Glutamate receptor subunit	832	–	—	12,653	n.e.
C10582	P07108	Acyl-CoA-binding protein (ACBP)	1	+	15,098	15,109ᶜ	15
G12582	P48735	Isocitrate dehydrogenase (NADP), mitochondrial precursor (EC 11142)	1	+	55,802	55,832	50
G14582	P39023	60S Ribosomal protein L3	225	–	—	8,904	10
G16582	P15880	40S Ribosomal protein S2 (S4)	1	+	34,328	34,302	35
C20582	P54198	HIRA protein	383	+	72,278	72,316ᶜ	65
I02582	P36404	ADP-ribosylation factor-like protein 2	1	–	—	12,703	12
I04582	P30086	Phosphatidylethanolamine-binding protein	1	+	27,046	27,029ᶜ	29
I06582	P25111	40S Ribosomal protein S25	1	+	17,678	17,685	23
I10582	P39023	60S Ribosomal protein L3	1	+	49,026	49,037ᶜ	50
I12582	P15880	40S Ribosomal protein S2 (S4)	1	+	34,328	34,329ᶜ	35
I14582	Q13098	G Protein pathway suppressor 1 (CPS1 protein)	193	–	—	17,996	18
I18582	P05092	Peptidyl-prolyl cis-trans isomerase A	1	+	21,441	21,460	23
I20582	P02570	Actin, cytoplasmic 1 (β-actin)	1	+	47,297	47,338	45
I24582	P23396	40S Ribosomal protein S3	1	+	29,749	29,761ᶜ	32
K04582	Q03827	Transcription factor ETR101	97	+	16,184	16,182	23
K08582	Q00403	Transcription initiation factor IIB (TFIIB)	1	+	38,133	38,156	38
K10582	O15143	ARP2/3 complex 41 kDa subunit (P41-ARC)	1	+	46,403	46,405	45
K12582	Q15666	Asparagine synthetase (fragment)	1	–	—	17,806	21
K14582	P49241	40S Ribosomal protein S3A	1	+	33,094	33,095ᶜ	35
K16582	Q99719	Cell division control-related protein	1	–	—	8,509	12
K18582	P04687	Tubulin α-1 chain	306	+	19,199	19,202	21
K20582	O00240	Dihydropyrimidinase-related protein-4 (DRP-4)	1	–	—	15,257	14
M02582	Q13885	β-Tubulin	253	–	—	26,301	27
M04582	P49368	T-complex protein 1, γ subunit (TCP-1-γ)	19	+	61,326	61,352	60
M10582	P02571	Actin, cytoplasmic 2 (γ-actin)	1	+	46,718	46,749	45
M12582	P32969	60S Ribosomal protein L9	1	–	—	9,950	14
M18582	Q13885	β Tubulin	1	+	54,291	54,307	58
M20582	P02768	Serum albumin precursor	116	+	59,076	59,116	60
M22582	P02570	Actin, cytoplasmic 1 (β-ACTIN)	1	+	47,297	47,306	45
M24582	Q06830	Thioredoxin peroxidase 2	1	+	26,231	26,222	25
O02582	Q02878	60S Ribosomal protein L6	1	+	35,457	35,460ᶜ	35
O04582	Q02878	60S Ribosomal protein L6	1	+	35,457	35,467ᶜ	35
O06582	P17080	CTP-binding nuclear protein RAN (TC4)	1	+	28,494	28,516ᶜ	30

TABLE 1—*Continued*

Clone MPMGp 800...	SP-TrEMBL database match		First matched amino acid in database sequence	Reading frame (RF)	Predicted protein size (kDa)	Expressed protein size, measured by MALDI-TOF-MS (kDa)	Expressed protein size, estimated by SDS-PAGE (kDa)*
	Accession No.	Protein name					
O08582	Q08379	Golgin-95	414	+	26,312	26,303	30
O10582	P01922	Hemoglobin α-chain	1	+	15,126	15,122	20
O14582	P21810	Bone/cartilage proteoglycan I precursor (biglycan) (PG-S1)	1	–	—	9,231	14
O16582	Q15597	Translation INITIATION FACTOR EIF-4γ (fragment)	215	+	58,069	58,108	60
O18582	Q14257	Calcium-binding protein ERC-55 precursor	24	+	42,157	42,165	60
O20582	Q02543	60S Ribosomal protein L18A	1	+	24,374	24,379ᶜ	26

Note. RF ± reading frame of insert in relation to His₆-tag. Expressed protein size was determined by MALDI-TOF-MS and estimated by SDS–PAGE referring to the band of the largest size visible against the *E. coli* background.

* n.e., no expression observed.
ᵇ Prior to reduction with DTT, abundant signals corresponding to the monomer, monomer + 1 glutathione residue, and the protein dimer were observed.
ᶜ Prior to reduction with DTT the determined mass was 295–315 Da higher (+ glutathione).

while in the remaining four clones (E18582, K10582, O02582, and O04582) no expression products could be detected (Table 1). His₆-tagged proteins were affinity-purified under denaturing conditions using Ni-NTA agarose beads and filter plates (Fig. 1). Six expression products, including E18582, K10582, O02582, and O04582, which were not detected in whole cellular protein extracts, could be identified after purification. Protein sizes were determined by SDS–PAGE and MALDI–TOF-MS (Fig. 2), and both data sets were
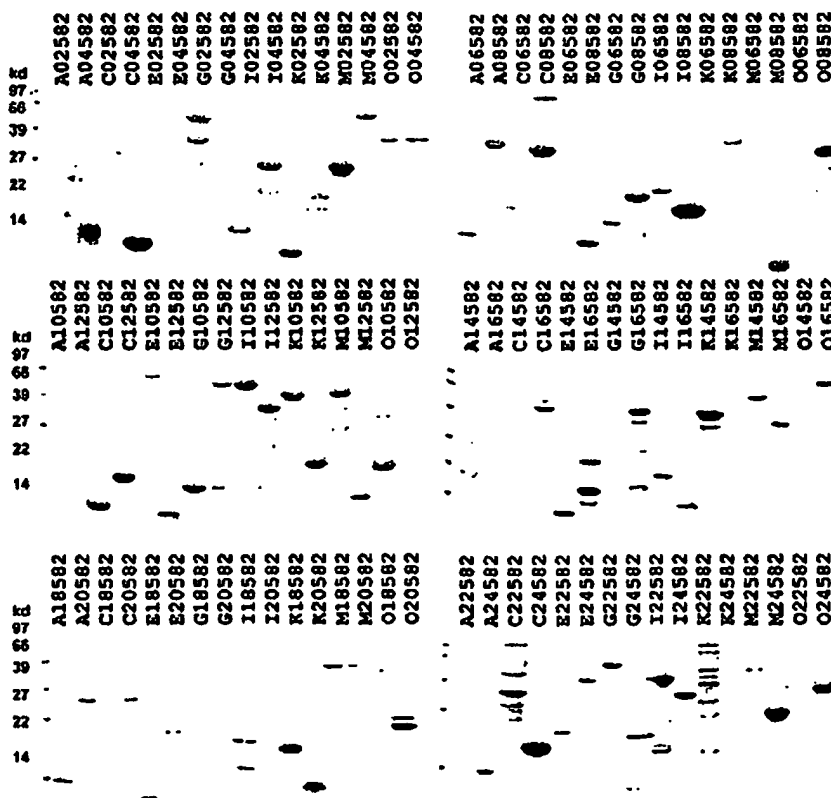

FIG. 1. SDS–PAGE of nickel chelate purified proteins. Following protein expression in microtiter plates, cells were lysed under denaturing conditions. His₆-tagged proteins were purified by nickel chelate chromatography in microtiter plates and were analyzed by SDS–PAGE, followed by Coomassie staining. Lanes are labeled using RZPD clone names, omitting the prefix MPMGp800.
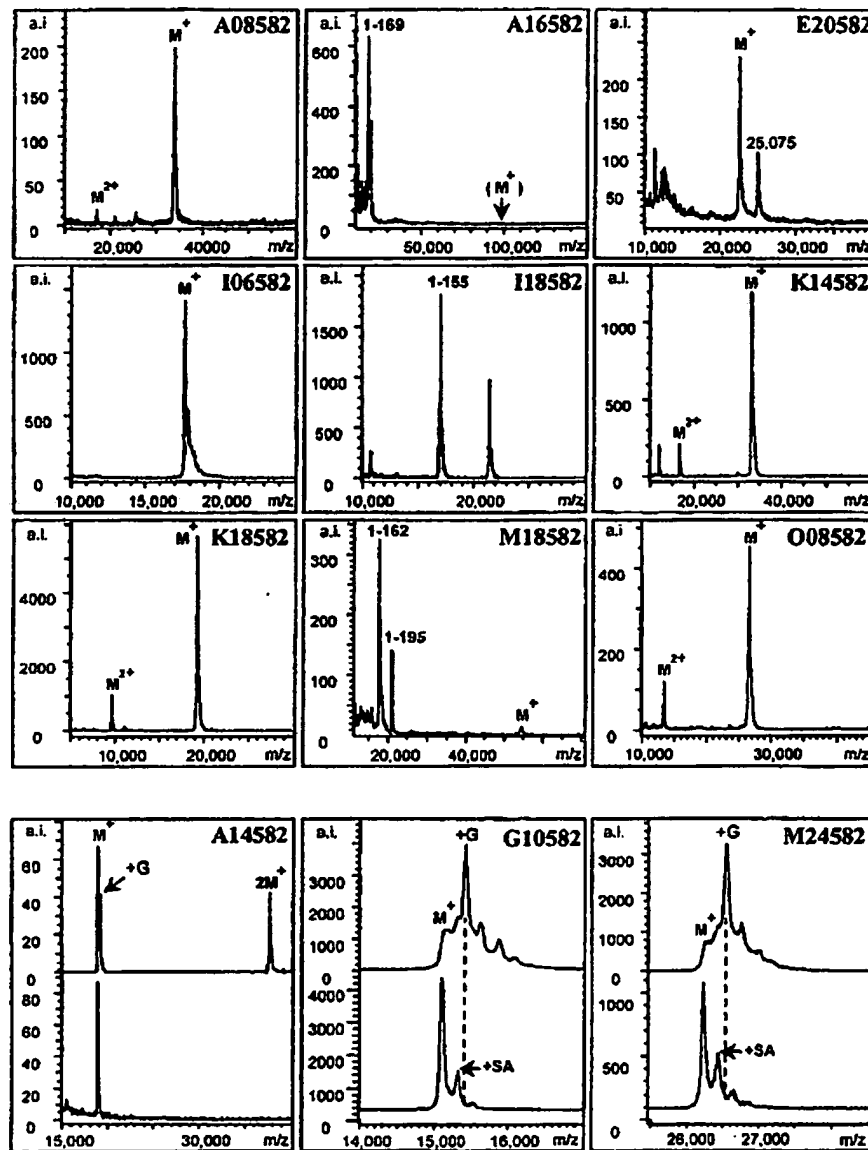
FIG. 2. MALDI-TOF-MS of nickel chelate purified proteins. Following nickel chelate affinity purification, the obtained eluates were analyzed by MALDI-TOF-MS. (Top three panels) Some recorded mass spectra; clones and labeling as in Fig. 1. M⁺ and M²⁺, singly and doubly charged molecular ions of expected expression products. The numbers of the first and last amino acids indicate assigned C-terminally truncated protein sequences. For clone A16582, the expected protein (98.098 kDa) could not be detected. For some expression products, the determined molecular masses exceeded the predicted values by approximately 300 Da (Table 1), indicative of glutathionylation. In addition, for A14582 a strong protein-dimer molecular ion signal was recorded indicative of protein-protein disulfide bridges. These indications were verified by reduction with DTT prior to the mass spectrometric analysis. (Bottom) Mass spectra obtained from A14582, G10582, and M24582 before (top spectrum) and after (bottom spectrum) reduction with DTT. +G, single glutathionylation; +SA, adduction of one sinapic acid molecule used as MALDI matrix.

compared to the corresponding values predicted from DNA sequencing data (Table 1).

As expected, the predicted molecular masses were considerably better matched by the masses determined with MALDI-TOF-MS than by those estimated from SDS-PAGE (Table 1). For most clones, the determined molecular mass deviated less than 0.1% from the pre-

dicted value. The fusion protein expected for clone A16582 (98.098 kDa) could not be detected; instead an abundant signal at m/z 18,446 dominated the recorded spectrum. Considering that the N-terminal sequence RGS-His₆ is vital for the applied affinity purification and that N-terminal methionine is usually not removed within E. coli if followed by arginine (Sherman

et al., 1985), this signal could be assigned to the truncated sequence 1–169 (expected *m/z* 18,442). SDS–PAGE of the purified A16582 protein showed a number of bands ranging from approximately 20 to approximately 100 kDa (Fig. 1). These results suggest that the protein is unstable in *E. coli.* An incomplete expression product was also observed for clone E18582. Various other clones produced abundant C-terminally truncated expression products in addition to the expected product. Figure 2 shows a selection of the recorded mass spectra including signal interpretation. Clone E20582 expressed a 25,075-Da protein of unknown identity in addition to the expected 22,579-Da protein. A possible explanation would be a frameshift mutation leading to a larger expression product in a subpopulation of the E20582 *E. coli* cells.

For the clones A14582, G10582, G12582, G16582, G20582, I04582. I10582. I12582. I24582. K04582. K10582, M24582. O02582. O04582, and O06582, the determined molecular masses exceeded the expected values by 0.290–0.320 kDa (three examples are shown in Fig. 2). This deviation is indicative of glutathionylation (attachment of one glutathione residue to a cysteine residue by formation of a disulfide bridge), an essential intermediate reaction of disulfide bridge reduction in *E. coli.* In addition, for the clone A14582 a strong signal corresponding to the molecular mass of the protein dimer was detected (Fig. 2). In addition to singly charged molecular ions, to a lower degree nonspecific multimers as well as doubly and triply charged molecular ions are formed during MALDI–TOF-MS of protein samples. Therefore, peak intensities must be taken into consideration to recognize protein dimers in the sample. Since the MALDI sample preparation conditions used (pH <2, 35% acetonitrile) denature most protein–protein interactions, a covalent linkage is likely to account for the observed dimers. Both glutathionylation and protein–protein disulfide bridges were verified by reduction with DTT prior to the mass spectrometric analysis. After reduction, the determined protein masses matched the expected masses in all cases within 0.1% maximum deviation, and no more protein dimers were detected (Fig. 2).

Expression products from inserts cloned in incorrect reading frames (RF–) were generally smaller than those from inserts in the correct reading frame (RF +, Table 1). As shown in Fig. 3, the molecular mass of expression products is correlated to the reading frame of the cDNA insert. Sixteen of 17 expression product smaller than 15 kDa derived from RF– clones, while 31 of 32 expression products of at least 20 kDa size derived from RF+ clones. Thus the molecular mass of a clone's expression product can be used as a measure to predict the reading frame of its cDNA insert, if the DNA sequence is unknown.

The screening of commonly available cDNA expression libraries for functional activities is complicated by large numbers of clones that do not express their cDNA inserts as proteins. By arraying, antibody screening,
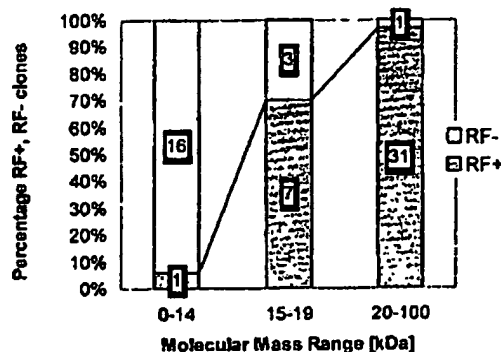

FIG. 3. Expression product size and reading frame. Relationship between the size of expressed recombinant protein and reading frame in the same clones as in Table 1. Clones represented by gray bars contain cDNA inserts translated in the correct reading frame (RF+), whereas in the other clones (white bars) translation can occur only in an incorrect reading frame (RF–). Numbers of clones are indicated in the bars.

and molecular mass detection, putative expression clones can be detected with a high level of efficiency. Therefore smaller numbers of clones must be assayed, and functional screening in microtiter plates becomes practicable. As an example, a GAPDH activity assay was adapted to the microtiter plate format. A positive control clone (D215) expressing human GAPDH as an RGS–His$_6$ tag fusion protein was introduced in exchange for one of the 96 hEx1 clones of Fig. 1. Protein expression was induced, and cells were lysed in 150 μl lysis buffer under nondenaturing conditions. GAPDH activities were measured in 0.1-μl aliquots of the lysates, and the positive control clone (D215) was clearly identified (Fig. 4). Duplicate experiments gave identical activity patterns with at least three additional clones (C04582, C06582, and C08582) above an arbitrary background. Two of these clones express products that did not match human proteins in the database, while the third represents a short out-of-frame fragment.

DISCUSSION

Structural genomics is expected to provide a link between DNA sequence information and protein function (Gaasterland, 1998; Kim, 1998; Rost, 1998). This requires the expression and characterization of large numbers of human proteins. We have shown that desired protein expression clones can be selected at high throughput from an arrayed cDNA library using a multistep screening procedure. This highly parallel approach seems to be an efficient alternative to the subcloning of individual cDNA sequences. In a first step, high-density protein filters are screened for putative expression clones (Büssow *et al.,* 1998; Lueking *et al.,* 1999). Those clones are then rearrayed into a sublibrary enriched for in-frame inserts. DNA sequence analysis of 93 randomly chosen clones of the hEx1
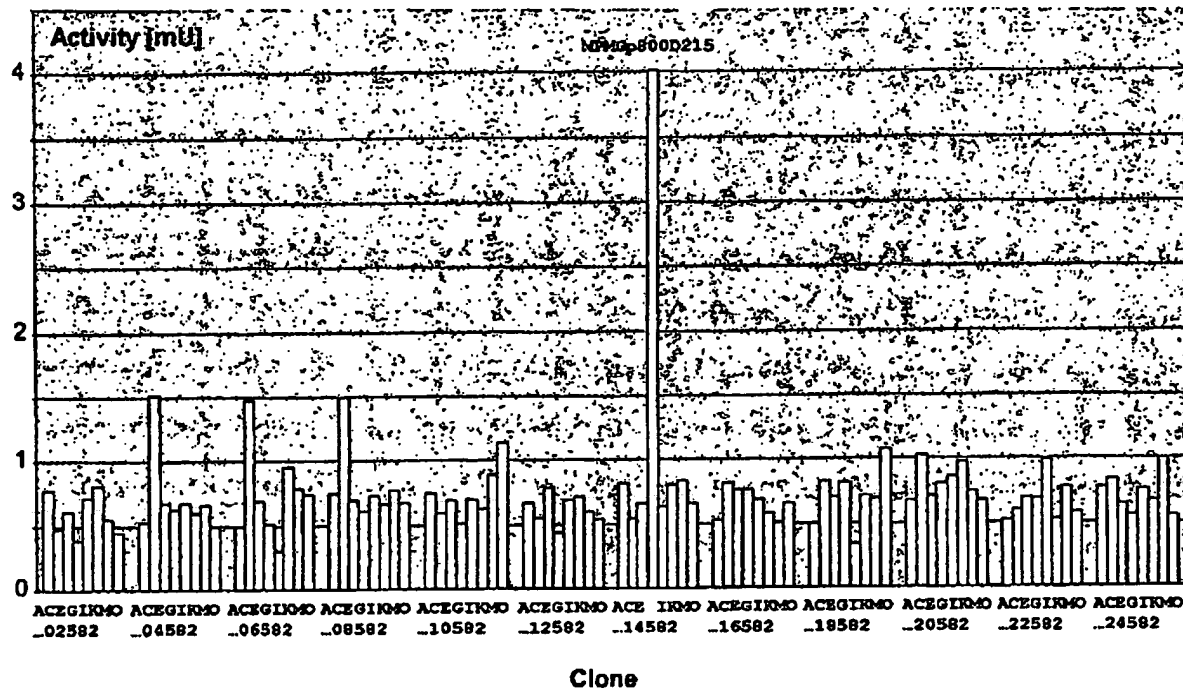
FIG. 4. GAPDH activity screen. Screening for GAPDH activity in bacterial lysates of 96 clones of the hExl library; clone numbers and labeling as in Fig. 1: positive control, MPMGp800D215 containing a GAPDH cDNA insert.

library showed that among the known genes, two-thirds (66%) expressed their inserts in the correct reading frame, reflecting the general efficiency of the screening method. The two-thirds (64%) of clones containing the complete coding sequence of a human protein do of course reflect the usual bias toward smaller products of cDNA libraries (Table 1).

In a second step, proteins are expressed in microtiter plates, and SDS–PAGE, nickel chelate affinity purification, and MALDI–TOF-MS are used to identify the best protein-expressing clones. DNA sequencing was used to correlate reading frames with protein sizes determined by SDS–PAGE and MALDI–TOF-MS. Although both methods returned compatible results, the latter was found to give the most accurate indication for the reading frame of cDNA inserts, because nearly all clones with expression products of at least 20 kDa had inserts in the correct reading frame (Fig. 3). This indicates that MALDI–TOF-MS size selection is a useful criterion for confirmation of expression clones. The molecular masses determined by MALDI–TOF-MS are in good agreement with the values predicted from the corresponding DNA sequences (Table 1). This reflects the expected experimental mass accuracy considering that all samples were analyzed using identical instrumental settings and the same external calibration constants. These examples demonstrate the power of MALDI–TOF-MS for characterizing cDNA expression products at high throughput. The detection is sensitive (midfemtomolar range), is rapid (<1 min per sample), and provides detailed and accurate information about

the status and purity of the expression products. Questions as to whether a certain clone produces high-quality protein for X-ray or NMR analysis or whether size exclusion chromatography following affinity purification can provide the necessary purity and homogeneity can be addressed early and at low cost. MALDI–TOF-MS analysis of whole libraries will provide a catalogue of expression clones for large numbers of human proteins. By including tryptic peptide mass fingerprinting, new clones will be directly identified by database comparison (Pappin et al., 1993).

In a third step, the microtiter plate technology was extended to functional screening. A spectrophotometric enzyme assay was developed that detects GAPDH expression clones among 96 hExl library clones, using nondenaturing bacterial lysates. It is expected that this kind of assay can be adapted to screen expression libraries for other biological activities in the microtiter plate format. Only 1/1000 of the bacterial lysate was used for the GAPDH assay. Therefore, the amount of protein is not expected to be limiting in future assays. Even if the bulk of a protein of interest is expressed in insoluble form, a small soluble fraction could be sufficient for detection. If necessary, affinity purification in microtiter plates can be used to reduce the background of E. coli proteins.

Up-scaling of the assay from 96 clones to the whole library will enable the detection of more clones with low to medium activity. This might include a certain degree of false-positive background but will also detect

new biological activity of yet uncharacterized human proteins.

In summary, the use of robot technology for handling and arraying of cDNA libraries, in combination with high-throughput microtiter plate techniques and MALDI-TOF-MS for the analysis of gene products, enables the generation of a catalogue of expression clones as a tool for the characterization of the human proteome.

ACKNOWLEDGMENTS

REFERENCES

Altschul, S. F., Gish, W., Miller, W., Myers. E. W., and Lipman, D. J. (1990). Basic local alignment search tool. *J. Mol. Biol.* 215: 403–410.

Aruffo, A. (1997). Transient expression of proteins using COS cells. In "Current Protocols in Molecular Biology" (F. M. Ausubel, R. Brent, R. E. Kingston, D. D. Moore, J. G. Seidman, et al., Eds.), pp. 16.13.11–16.13.17. Wiley, New York.

Bairoch, A., and Apweiler, R. (1998). The SWISS-PROT protein sequence data bank and its supplement TrEMBL in 1998. *Nucleic Acids Res.* 26: 38–42.

Buckholz, R. G., and Gleeson, M. A. (1991). Yeast systems for the commercial production of heterologous proteins. *Bio/Technology* 9: 1067–1072.

Büssow, K., Cahill, D., Nietfeld, W., Bancroft, D., Scherzinger, E., Lehrach, H., and Walter, G. (1998). A method for global protein expression and antibody screening on high-density filters of an arrayed cDNA library. *Nucleic Acids Res.* 26: 5007–5008.

Gaasterland, T. (1998). Structural genomics: Bioinformatics in the driver's seat. *Nat. Biotechnol.* 16: 625–627.

Gottesman, S. (1996). Proteases and their targets in *Escherichia coli. Annu. Rev. Genet.* 30: 465–506.

Gubler, U., and Hoffman, B. J. (1983). A simple and very efficient method for generating cDNA libraries. *Gene* 25: 263–269.

Heinz, F., and Freimüller, B. (1982). Glyceraldehyde-3-phosphate dehydrogenase from human tissues. *Methods Enzymol.* 89: 301–305.

Hockney, R. C. (1994). Recent developments in heterologous protein production in *Escherichia coli. Trends Biotechnol.* 12: 456–463.

Hoheisel, J. D., Lennon, G., Zehetner, G., and Lehrach, H. (1991). Use of high coverage reference libraries of *Drosophila melanogaster* for relational data analysis—A step towards mapping and sequencing of the genome. *J. Mol. Biol.* 220: 903–914.

Hwang, K. Y., Chung, J. H., Kim, S.-H., Han, Y. S., and Cho, Y. (1999). Structure-based identification of a novel NTPase from *Methanococcus jannaschii. Nat. Struct. Biol.* 6: 691–696.

Kim, S. H. (1998). Shining a light on structural genomics. *Nat. Struct. Biol.* 5(Suppl.): 643–645.

Kingston, R. E., Kaufman, R. J., Bebbington, C. R., and Rolfe, M. R. (1997). Amplification using CHO cell expression vectors. In "Current Protocols in Molecular Biology" (F. M. Ausubel, R. Brent, R. E. Kingston, D. D. Moore, J. G. Seidman, et al., Eds.), p. 16.14.11. Wiley, New York.

Lehrach, H., Bancroft, D., and Maier, E. (1997). Robotics, computing, and biology. *Interdisc. Sci. Rev.* 22: 37–44.

Lueking, A., Horn, M., Eickhoff, H., Büssow, K., Lehrach, H., and Walter, G. (1999). Protein microarrays for gene expression and antibody screening. *Anal. Biochem.* 270: 103–111.

Maier, E., Bancroft, D. R., and Lehrach, H. (1997). Large-scale library characterization. In "Automation Technologies for Genome Characterization" (T. J. Beugelsdijk, Ed.), pp. 65–88. Wiley, New York.

Maier, E., Meier-Ewert, S., Ahmadi, A. R., Curtis, J., and Lehrach, H. (1994). Application of robotic technology to automated sequence fingerprint analysis by oligonucleotide hybridisation. *J. Biotechnol.* 35: 191–203.

Makrides, S. C. (1996). Strategies for achieving high-level expression of genes in *Escherichia coli. Microbiol. Rev.* 60: 512–538.

Miller, L. K. (1993). Baculoviruses: High-level expression in insect cell. *Curr. Opin. Genet. Dev.* 3: 97–101.

Pappin, D. J. C., Hojrup, P., and Bleasby, A. J. (1993). Rapid identification of proteins by peptide-mass fingerprinting. *Curr. Biol.* 3: 327–332.

Rost, B. (1998). Marrying structure and genomics. *Structure* 6: 259–263.

Sherman, F., Stewart, J. W., and Tsunasawa, S. (1985). Methionine or not methionine at the beginning of a protein. *BioEssays* 3: 27–31.

Zarembinski, T. I., Hung, L. W., Mueller-Dieckmann, H. J., Kim, K. K., Yokota, H., Kim, R., and Kim, S. H. (1998). Structure-based assignment of the biochemical function of a hypothetical protein: A test case of structural genomics. *Proc. Natl. Acad. Sci. USA* 95: 15189–15193.